Visualizing Black Holes With a Physically Accurate Relativistic Ray Marcher

Pontus Rosqvist

Lund University*

Abstract

The theory for manifolds and geodesics on manifolds is informally presented and the geodesic equation is derived. Geodesics are interpreted to be, and indeed are, the paths light travels along in the space-time manifold. Adaptive Runge-Kutta methods are applied to solving the geodesic equation and images of a black hole are rendered where rays are sent from a camera and follow geodesic paths under the influence of a non-rotating black hole.

Introduction

This project aims to create a relativistic ray marcher which renders a scene where the light is bent by the metric induced by a non-rotating black hole. To this end we need to construct rays that bend correctly as we step along them, this can be done if we solve the geodesic equation:

$$\ddot{\gamma}^k(t) + \sum_{i,j=1}^{n} \Gamma_{ij}^k(\gamma(t))\dot{\gamma}^i(t)\dot{\gamma}^j(t) = 0 \qquad 1 \leq k \leq n \tag{1}$$

where $\gamma : \mathbb{R} \to \mathbb{R}^n$ is a path in space and $\Gamma_{ij}^k : \mathbb{R}^n \to \mathbb{R}$ are the Christoffel symbols which are a set of functions which in a sense describe how the ambient space influences "straight" paths. The Christoffel symbols are intrinsic functions associated with the ambient space we are considering and for the usual Euclidean, i.e. flat, space we have that the Christoffel symbols are all zero.

The geodesic equation generalises the notion of "straight" paths for curved spaces in the sense that a geodesic is the path that a particle (or a photon) would follow if it is set in motion from a point in some direction without any forces acting on it as it moves. Since the Christoffel symbols are identically zero for Euclidean space, $\mathbb{R}^n$, the geodesic equation becomes

$$\ddot{\gamma}^k(t) = 0$$

and the solution to this differential equation is

$$\gamma(t) = at + b$$

for some constant vectors $a, b \in \mathbb{R}^n$. This is a straight line which we except a particle to follow if it is set in motion and no external forces affect it.

To understand the geodesic equation it is necessary to talk about Riemannian manifolds and the Levi-Civita connection. We will introduce this later but for now we note that the Levi-Civita connection is an operator that acts on vector fields similarly to a derviative, with the Levi-Civita connection the geodesic equation simplifies to the following

$$\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t) = 0$$

Another interesting differential equation that we will encounter later is the Jacobi equation which is the following differential equation

$$\nabla_{\dot{\gamma}(t)}\nabla_{\dot{\gamma}(t)}J(\gamma(t)) + R(J(\gamma(t)), \dot{\gamma}(t))\dot{\gamma}(t) = 0$$

a vector field, J, that satisfies the above equation is called a Jacobi field which measures the divergence of geodesics.

*email:pontus.rosqvist@gmail.com

Method & Theory

For an excellent introduction into the beautiful theory of Riemannian manifolds we refer the reader to the textbook by Gudmundsson titled *An Introduction to Riemannian Geometry* [1].

Manifolds

Consider the line that was defined previously

$$\gamma(t) = at + b$$

where $a, b \in \mathbb{R}^n$, clearly it does not matter which dimension, i.e. the value of n, we choose to use here. The line is always, in a sense, one-dimensional and if we bend and twist the path we do not change this property. If we go up one dimension we can define a surface in the following way

$$\Gamma(s,t) = \begin{bmatrix} s \\ t \\ at^2 + bs^2 + c \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^n$$

where $a, b, c \in \mathbb{R}$. Even if we bend the surface it is still intrinsically two-dimensional, Similarly we might define higher-dimensional analogues of curves and surfaces but regardless of the dimension of the ambient space they are defined in ($\mathbb{R}^n$ in the two previous examples) they have their own intrinsic dimension and structure. Manifolds are a mathematical structure which generalises this notion of curves and surfaces and formalises what it means for a manifold to be continuous and even non-flat and this definition does not require defining the manifold in an ambient space. This is an important distinction when it comes to mathematical rigour but will not affect us.

For future reference note that the set of all smooth functions defined on a manifold, M, is denoted by $C^\infty(M)$.

Vector Fields and The Tangent Space

Consider the two dimensional manifold defined above embedded in the three-dimensional space $\mathbb{R}^3$:

$$\Gamma(s,t) = \begin{bmatrix} s \\ t \\ at^2 + bs^2 + c \end{bmatrix},$$

it is important to note that the manifold is the *image* of the function and the function has no relation to the manifold otherwise. At every point of a manifold we can define a tangent space which, informally corresponds to the set of all possible tangents for any path through that point. For our example above a basis for the tangent space can be defined in the following way:

$$\frac{\partial \Gamma}{\partial s} = \begin{bmatrix} 1 \\ 0 \\ 2bs \end{bmatrix} \qquad \frac{\partial \Gamma}{\partial t} = \begin{bmatrix} 0 \\ 1 \\ 2at \end{bmatrix}$$

by convention (and other more nebulous reasons) the Γ is usually dropped and the tangent space is seen as directional derivatives. If we denote a manifold by M then it's tangent space at a point $p \in M$ is denoted by T_pM.

Once we have defined a tangent space at every point we would like to define vector fields. Vector fields associate each point from the manifold to a tangent in the tangent space, i.e. if X is a vector field then for $p \in M$ we get $X(p) \in T_pM$. Since a vector field is not a function in the usual sense (points are mapped to different domains) we usually write X_p for the tangent vector at p instead.

Since manifolds are continuous it is possible to define what it means for a vector field to be continuous, this is quite technical but is intuitively easy to understand, consider the vector field defined on the unit circle in figure 1 below.

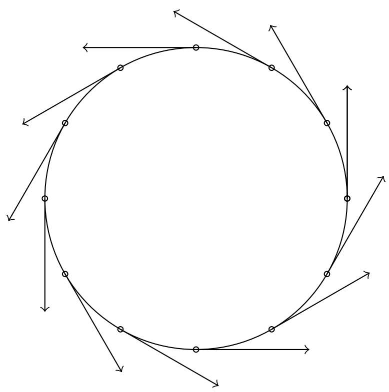

Figure 1: Some points of a continuous vector field defined on the unit circle.

The vector field in figure 1 has the following formula

$$X_p = \begin{bmatrix} -y \\ x \end{bmatrix}, \qquad \text{where } p = \begin{bmatrix} x \\ y \end{bmatrix}$$

Clearly the unit circle is a one-dimensional manifold and at each point we have defined a vector tangent to the unit circle and between each point the vectors are continuously defined. Since tangent vectors are directional derivatives (along that vector at that point) a vector field can be thought of as acting on functions on the whole manifold. The set of smooth vector fields on a manifold is denoted by $C^\infty(TM)$.

Riemannian Manifolds

A Riemannian manifold is a manifold with a metric defined on it. A metric is, in essence, a continuous (and separate) inner product for the tangent space at each point of a manifold, this is exactly what makes manifolds curved. Note that any inner product defined on a vector space V, $(\cdot,\cdot) : V \times V \to V$ maps vectors to numbers with the following properties

$$\begin{aligned} (X, X) &\geq 0, \\ (X, X) &= 0 \quad \iff \quad X = 0, \\ (X, Y) &= (Y, X), \\ (X + Y, Z) &= (X, Z) + (Y, Z). \end{aligned}$$

Given any basis of the vector space one can find a matrix, associated with the inner product such that

$$(X, Y) = \begin{bmatrix} X_1 & \dots & X_n \end{bmatrix} \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix} \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$$

a Riemannian metric on a manifold is then just an inner product on each tangent space of that manifold in a smooth sense, if we have smooth vector fields which form a basis of the tangent space of the manifold at each point then we can find the values of the matrix of the Riemannian metric at each point

$$g_p(\cdot,\cdot) = \begin{bmatrix} g_{11}(p) & \dots & g_{1n}(p) \\ \vdots & \ddots & \vdots \\ g_{n1}(p) & \dots & g_{nn}(p) \end{bmatrix}$$

this basis is always possible to construct locally but not all manifolds admit set of vector fields which form a basis globally.

The Levi-Civita Connection

The Levi-Civita connection, $\nabla : C^\infty(TM) \times C^\infty(TM) \to C^\infty(TM)$, is a fundamental operator that has similar properties to the usual directional derivative for manifolds which acts on vector fields to give a new vector field. The Levi-Civita Connection has the following properties, let $X, Y, Z \in C^\infty(TM)$, $\lambda, \mu \in \mathbb{R}$ and $f, g \in C^\infty(M)$ then

$$\begin{aligned} \nabla_X(\lambda Y + \mu Z) &= \lambda \nabla_X Y + \mu \nabla_X Z \\ \nabla_X(f \cdot Y) &= X(f) \cdot Y + f \cdot \nabla_X Y \\ \nabla_{(f \cdot X + g \cdot Y)} Z &= f \cdot \nabla_X Z + g \cdot \nabla_Y Z \end{aligned}$$

any operator which satisfies these relations is called a connection and many of these exist, *the* (unique) Levi-Civita connection also satisfies the following relation

$$X(g(Y, Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z)$$

If we let $\left\{ \frac{\partial}{\partial x_i} \right\}_{i=1}^n$ be a set of vector fields on a manifold which form an orthonormal basis of the tangent space at each point on the manifold then the Christoffel symbols $\Gamma_{ij}^k : M \to \mathbb{R}$ associated with the Levi-Civita connection are defined in the following way

$$\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j} = \sum_{k=1}^n \Gamma_{ij}^k \frac{\partial}{\partial x_k}.$$

The Christoffel symbols can be defined in terms of the Riemannian metric on the manifold, we will not derive it here but the formula is the following:

$$\Gamma_{ij}^k = \frac{1}{2} \sum_{l=1}^n g^{kl} \left(\frac{\partial g_{jl}}{\partial x_i} + \frac{\partial g_{li}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_l} \right)$$

Where g^{kl} is element kl of the inverse of the metric. The Levi-Civita connection is also called the covariant derivative and the interpretation is that it computes the orthogonal projection of the directional derivative of vector fields along the tangent space.

The Geodesic Equation

Geodesics are paths defined on a manifold, $\gamma : \mathbb{R} \to M$, and are solutions to the following differential equation

$$\nabla_{\dot\gamma(t)} \dot\gamma(t) = 0.$$

This form of the formula is, however, entirely useless when it comes to actually finding geodesics numerically on a manifold by solving the differential equation. We can however rewrite this into a more useful form, note that if we have a set of vector fields which forms an orthonormal basis at every point of the tangent space then the tangent of the path can be written in the following way

$$\dot\gamma(t) = \sum_{j=1}^n \dot\gamma_j(t) \left(\frac{\partial}{\partial x_j} \right)_{\gamma(t)}$$

then we can apply the previously defined properties of the Levi-Civita connection in the following way

$$\begin{aligned}
\nabla_{\dot\gamma(t)}\dot\gamma(t) &= \sum_{j=1}^{n} \nabla_{\dot\gamma(t)} \left(\dot\gamma_j(t) \left(\frac{\partial}{\partial x_j}\right)_{\gamma(t)} \right) \\
&= \sum_{j=1}^{n} \left(\dot\gamma(t)(\dot\gamma_j(t)) + \dot\gamma_j(t) \left(\nabla_{\dot\gamma(t)} \frac{\partial}{\partial x_j}\right)_{\gamma(t)} \right) \\
&= \sum_{j=1}^{n} \Bigg(\ddot\gamma_j(t) \left(\frac{\partial}{\partial x_j}\right)_{\gamma(t)} + \\
&\qquad\qquad \dot\gamma_j(t) \left(\nabla_{\sum_{i=1}^{n} \dot\gamma_i(t)\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j}\right)_{\gamma(t)} \Bigg) \\
&= \sum_{j=1}^{n} \Bigg(\ddot\gamma_j(t) \left(\frac{\partial}{\partial x_j}\right)_{\gamma(t)} + \\
&\qquad\qquad \sum_{i=1}^{n} \dot\gamma_i(t)\dot\gamma_j(t) \left(\nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j}\right)_{\gamma(t)} \Bigg) \\
&= \sum_{j=1}^{n} \Bigg(\ddot\gamma_j(t) \left(\frac{\partial}{\partial x_j}\right)_{\gamma(t)} + \\
&\qquad\qquad \sum_{i=1}^{n} \dot\gamma_i(t)\dot\gamma_j(t) \sum_{k=1}^{n} \Gamma_{ij}^{k}(\gamma(t)) \left(\frac{\partial}{\partial x_k}\right)_{\gamma(t)} \Bigg) \\
&= \sum_{k=1}^{n} \left(\ddot\gamma_k(t) + \sum_{i=1}^{n}\sum_{j=1}^{n} \dot\gamma_i(t)\dot\gamma_j(t)\Gamma_{ij}^{k}(\gamma(t)) \right) \left(\frac{\partial}{\partial x_k}\right)_{\gamma(t)}
\end{aligned}$$

for this vector field to be zero each coordinate of the resulting vector field has to be zero for all t which means that we get a set of n non linear differential equations of the following form

$$\ddot\gamma_k(t) + \sum_{i=1}^{n}\sum_{j=1}^{n} \dot\gamma_i(t)\dot\gamma_j(t)\Gamma_{ij}^{k}(\gamma(t)) = 0$$

which is the geodesic equation.

The Jacobi Equation

The Jacobi equation is the following differential equation

$$\nabla_{\dot\gamma(t)}\nabla_{\dot\gamma(t)}J(\gamma(t)) + R(J(\gamma(t)), \dot\gamma(t))\dot\gamma(t) = 0.$$

The interpretation of this differential equation is that the vector field J, which is defined along a geodesic, measures how much infinitesimally close geodesics will diverge from the current one. An example of a Jacobi field along a geodesic can be seen in figure 2 on the unit sphere in $\mathbb{R}^3$.

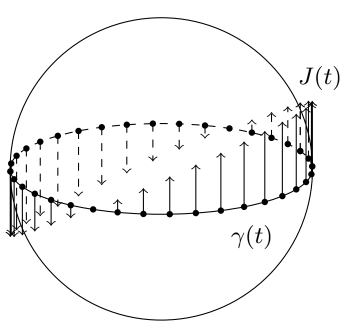

Figure 2: The great circle along the equator of the sphere is a geodesic and the attached vector field is a Jacobi field.

If we, for a second, remember that we would like to image a black hole we could leverage this interpretation by solving this equation along pilot geodesics and recording the maximal separation at each geodesic. We would then be able to quantify how much geodesics spread apart when sent from a pixel which could give us an adaptive supersampling where we make sure to send many rays for a pixel with a high maximal separation and few rays for a pixel with a low maximal separation to ensure that we hit the scene evenly with rays.

This form of the differential equation is, again, useless for actually solving the Jacobi equation numerically and it even depends on the Riemannian curvature of the manifold which is calculated from the Christoffel symbols and their derivatives which adds another layer of complexity.

The derivation of the formula for the Jacobi equation which is more friendly for numerical computations can be found on page 9 in the appendix. The formula is the following:

$$\begin{aligned}
0 = \ddot J_k + 2\dot J^{\mathrm{T}}\Gamma^k J + J^{\mathrm{T}} \left(\sum_m \dot\gamma^m \frac{\partial}{\partial x_m}\Gamma^k \right) J \\
+ \dot J^{\mathrm{T}}\Gamma^k\dot\gamma + \left(\Gamma^k\dot\gamma\right)\sum_m \left(J^{\mathrm{T}}\Gamma^m J\right) + \sum_l \dot\gamma^l \left(J^{\mathrm{T}} R_l^k \dot\gamma\right)
\end{aligned}$$

To make the formula easier to write down we have defined the following symbols

$$\begin{aligned}
\Gamma^k &= \begin{bmatrix} \Gamma_{11}^k & \dots & \Gamma_{1n}^k \\ \vdots & \ddots & \vdots \\ \Gamma_{n1}^k & \dots & \Gamma_{nn}^k \end{bmatrix} \\
R_l^k &= \begin{bmatrix} R_{l11}^k & \dots & R_{l1n}^k \\ \vdots & \ddots & \vdots \\ R_{ln1}^k & \dots & R_{lnn}^k \end{bmatrix} \\
e &= \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}
\end{aligned}$$

and this actually enables us to rewrite the geodesic equation into a quite elegant form

$$\ddot\gamma^k + \dot\gamma^{\mathrm{T}}\Gamma^k\dot\gamma = 0.$$

Adaptive Runge-Kutta Methods

It is now clear that to be able to find the path that photons will follow and how many photons we need to send at each pixel we need to be able to solve non-linear second order ordinary differential equations. Runge-Kutta methods are numerical methods for approximating solutions to non-linear first order initial value differential equations. These differential equations are of the form

$$\begin{aligned}
\dot y(t) &= f(t, y(t)) \\
y(0) &= x
\end{aligned}$$

to solve this numerically we need to restrict ourselves to approximating the solution on an interval, let this interval be $[0, T]$ and place N equally spaced points in this interval, we introduce the following symbols

$$\begin{aligned}
t_i &= \frac{T}{N-1} \cdot i = \Delta t \cdot i \\
y_i &= y(t_i)
\end{aligned}$$

the simplest Runge-Kutta method is the most natural explicit Euler which approximates the solution in the following way

$$\begin{aligned} y_{i+1} &= y_i + \Delta t f(t_i, y_i) \\ y_0 &= x \end{aligned}$$

For each step we take we will have to detect if the ray has collided with any object in the scene, thus we will need to take large steps to have any hope of rendering an image in a reasonable amount of time. As far as numerical integration methods go the explicit Euler is first order convergent which means that as the step size goes down the error decreases linearly.

A large step size will lead to large cumulative numerical errors especially close to the black hole where a small deviation will lead to a large difference in where the ray finally ends up. Thus the explicit Euler method will not be sufficient for solving our differential equation. A suitable method for solving is thus a Runge-Kutta method with a higher order of convergence. A general explicit Runge-Kutta method can be described by its Butcher tableau which is of the form [2]

$$\begin{array}{c|ccccc} 0 & & & & & \\ c_2 & a_{21} & & & & \\ c_3 & a_{31} & a_{32} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ c_s & a_{s1} & a_{s2} & \dots & a_{ss-1} & \\ \hline & b_1 & b_2 & \dots & b_{s-1} & b_s \end{array}$$

the update rule for a Runge-Kutta method with this Butcher tableau is the following

$$\begin{aligned} k_1 &= f(t_i, \quad y_i \quad) \\ k_2 &= f(t_i + c_2\Delta t, y_i + (a_{21}k_1 \quad)\Delta t) \\ k_3 &= f(t_i + c_3\Delta t, y_i + (a_{31}k_1 + a_{32}k_2 \quad)\Delta t) \\ &\vdots \\ k_s &= f(t_i + c_s\Delta t, y_i + (a_{s1}k_1 + \ldots + a_{ss-1}k_{s-1})\Delta t) \\ y_{i+1} &= y_i + \Delta t(b_1k_1 + \ldots + b_sk_s) \end{aligned}$$

By computing an update rule like this with well chosen parameters one can get a higher order of convergence and be able to take larger step sizes while still accurately solving the differential equation.

One problem with this, however, is that we do not want to use a fixed step size, when we know that the error will be high we want to use a small step size and when we know that the error will be low we can use a larger step size. This is the basis of adaptive methods which update the step size such that the error is always kept at some approximate low value. Since the true error is not known in general one usually approximates the error with an embedded method which has a lower order of convergence. These methods are described by an extended Butcher tableau

$$\begin{array}{c|ccccc} 0 & & & & & \\ c_2 & a_{21} & & & & \\ c_3 & a_{31} & a_{32} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ c_s & a_{s1} & a_{s2} & \dots & a_{ss-1} & \\ \hline & b_1 & b_2 & \dots & b_{s-1} & b_s \\ & b_1^* & b_2^* & \dots & b_{s-1}^* & b_s^* \end{array}$$

and the new step, y_{i+1}, as well as the error estimate, e_{i+1}, are computed in the following way

$$\begin{aligned} k_1 &= f(t_i, \quad y_i \quad) \\ k_2 &= f(t_i + c_2\Delta t, y_i + (a_{21}k_1 \quad)\Delta t) \\ k_3 &= f(t_i + c_3\Delta t, y_i + (a_{31}k_1 + a_{32}k_2 \quad)\Delta t) \\ &\vdots \\ k_s &= f(t_i + c_s\Delta t, y_i + (a_{s1}k_1 + \ldots + a_{ss-1}k_{s-1})\Delta t) \\ y_{i+1} &= y_i + \Delta t(b_1k_1 + \ldots + b_sk_s) \\ e_{i+1} &= y_{i+1} - y_{i+1}^* \quad = \Delta t \sum_{j=1}^{i-1} (b_j - b_j^*)k_j \end{aligned}$$

Given some minimal and maximal error tolerances , τ_1 and τ_2, and a desired error tolerance, τ, we can verify that the error lies inside the tolerance

$$\tau_1 \leq \|e_{i+1}\|_2 \leq \tau_2$$

The error estimate can then be used to update the step size with the following formula

$$\Delta t_{i+1} = \begin{cases} \Delta t_i, & \text{if } \tau_1 \leq \|e_{i+1}\|_2 \leq \tau_2 \\ 0.9\Delta t\sqrt{\frac{\tau}{2\|e_{i+1}\|_2}}, & \text{otherwise} \end{cases}$$

one can then either accept the step and move on with the new step size or reject the step and try again with the new step size.

The attentive reader will notice that these methods cannot be immediately applied to our problem since these methods can solve differential equations of the form

$$\begin{aligned} \dot{y}(t) &= f(t, y(t)) \\ y(0) &= x \end{aligned}$$

and both of the differential equations we would like to solve can be written on the form

$$\begin{aligned} \ddot{y}(t) &= f(t, y(t), \dot{y}(t)) \\ &\begin{cases} y(0) = x \\ \dot{y}(0) = v \end{cases} \end{aligned}$$

this is, however, not a problem since any second order differential equation with n equations can be rewritten as a first order differential equation with $2n$ equations in the following way

$$\begin{aligned} \dot{z}(t) &= f(t, z(t)) \\ z(0) &= z_0 \end{aligned}$$

where

$$\begin{aligned} z(t) &= \begin{bmatrix} y(t) \\ \dot{y}(t) \end{bmatrix} \\ g(t, z(t)) &= \begin{bmatrix} \dot{y}(t) \\ f(t, y(t), \dot{y}(t)) \end{bmatrix} \\ z_0 &= \begin{bmatrix} x \\ v \end{bmatrix} \end{aligned}$$

which enables us to use Runge-Kutta methods to approximate solutions to the geodesic equation.

Implementation

Deriving The Christoffel Symbols

To begin with we must define the manifold and choose a set of coordinates. Our Riemannian manifold is $\mathbb{R}^4$ with a special metric, the coordinates of a point p will be written in the following way

$$p = \begin{bmatrix} p^t & p^x & p^y & p^z \end{bmatrix}^{\mathrm{T}}$$

Where the first coordinate is the time at that point and the next three coordinates are the spatial coordinates.

We are working with a specific manifold which is the space-time manifold we live in, the metric on this manifold satisfies the Einstein field equations which is a differential equation. People smarter than us have already produced closed form solutions of this metric induced by a black hole positioned in the origin. The metric is the following:

$$g = \begin{bmatrix} -\frac{\left(1-\frac{r_s}{4R}\right)^2}{\left(1+\frac{r_s}{4R}\right)^2} & & & \\ & \left(1+\frac{r_s}{4R}\right)^4 & & \\ & & \left(1+\frac{r_s}{4R}\right)^4 & \\ & & & \left(1+\frac{r_s}{4R}\right)^4 \end{bmatrix}$$

previously we promised that the metric would be a matrix of functions defined on the manifold which defines an inner product on the tangent space at each point of the manifold. This proposed metric is however not positive definite which would be a requirement. To this we say that this is why general relativity studies pseudo-Riemannian manifolds and don't worry about it.

In the metric we define $R = \sqrt{x^2+y^2+z^2}$ and note that the metric does not depend on the time, t, this makes sense since we except light to bend the same way regardless of *when* it occured. We also see that the only spatial dependency of the metric is how close a point is to the origin, where the black hole is, which makes sense since we expect the bending of light to be rotationally symmetric around a black hole. Additionally note that there is a parameter here which is the event horizon $\frac{r_s}{4}$ which is a radius we have to choose inside of which light cannot escape.

To compute the Christoffel symbols with the least work needed we first note that the metric can be written in the following way

$$g = \begin{bmatrix} f_1 & & & \\ & f_2 & & \\ & & f_2 & \\ & & & f_2 \end{bmatrix}$$

$$f_1(x,y,z) = -\frac{\left(1-\frac{r_s}{4R}\right)^2}{\left(1+\frac{r_s}{4R}\right)^2}$$

$$f_2(x,y,z) = \left(1+\frac{r_s}{4R}\right)^4$$

To compute the Christoffel symbols we need to compute all partial derivative of these functions which are the following

$$\begin{aligned}
\frac{\partial f_1}{\partial t} &= 0, & \frac{\partial f_2}{\partial t} &= 0 \\
\frac{\partial f_1}{\partial x} &= -\frac{x r_s}{R^3}\frac{1-\frac{r_s}{4R}}{\left(1+\frac{r_s}{4R}\right)^3}, & \frac{\partial f_2}{\partial x} &= -\frac{x r_s}{R^3}\left(1+\frac{r_s}{4R}\right)^3 \\
\frac{\partial f_1}{\partial y} &= -\frac{y r_s}{R^3}\frac{1-\frac{r_s}{4R}}{\left(1+\frac{r_s}{4R}\right)^3}, & \frac{\partial f_2}{\partial y} &= -\frac{y r_s}{R^3}\left(1+\frac{r_s}{4R}\right)^3 \\
\frac{\partial f_1}{\partial z} &= -\frac{z r_s}{R^3}\frac{1-\frac{r_s}{4R}}{\left(1+\frac{r_s}{4R}\right)^3}, & \frac{\partial f_2}{\partial z} &= -\frac{z r_s}{R^3}\left(1+\frac{r_s}{4R}\right)^3
\end{aligned}$$

Before we continue by using the previously defined formula for the Christoffel symbols we need to find the inverse of the metric which is, since the matrix is diagonal, clearly the following

$$g^{-1} = \begin{bmatrix} \frac{1}{f_1} & & & \\ & \frac{1}{f_2} & & \\ & & \frac{1}{f_2} & \\ & & & \frac{1}{f_2} \end{bmatrix}$$

Remember that the formula for the Christoffel symbols are

$$\Gamma^k_{ij} = \frac{1}{2}\sum_{l=1}^{n} g^{kl}\left(\frac{\partial g_{jl}}{\partial x_i} + \frac{\partial g_{li}}{\partial x_j} - \frac{\partial g_{ij}}{\partial x_l}\right)$$

since the inverse matrix is diagonal the index l necessarily has to be equal to k, additionally note that the formula is symmetric if the indices i and j are swapped, thus we need to compute the following expressions:

$$\begin{aligned}
\Gamma^t_{tt} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{tt}}{\partial t} + \frac{\partial g_{tt}}{\partial t} - \frac{\partial g_{tt}}{\partial t}\right) &= 0 \\
\Gamma^t_{tx} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{xt}}{\partial t} + \frac{\partial g_{tt}}{\partial x} - \frac{\partial g_{tx}}{\partial t}\right) &= \frac{1}{2}g^{tt}\frac{\partial g_{tt}}{\partial x} \\
\Gamma^t_{ty} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{yt}}{\partial t} + \frac{\partial g_{tt}}{\partial y} - \frac{\partial g_{ty}}{\partial t}\right) &= \frac{1}{2}g^{tt}\frac{\partial g_{tt}}{\partial y} \\
\Gamma^t_{tz} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{zt}}{\partial t} + \frac{\partial g_{tt}}{\partial z} - \frac{\partial g_{tz}}{\partial t}\right) &= \frac{1}{2}g^{tt}\frac{\partial g_{tt}}{\partial z} \\
\Gamma^t_{xx} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{xt}}{\partial x} + \frac{\partial g_{tx}}{\partial x} - \frac{\partial g_{xx}}{\partial t}\right) &= 0 \\
\Gamma^t_{xy} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{yt}}{\partial x} + \frac{\partial g_{tx}}{\partial y} - \frac{\partial g_{xy}}{\partial t}\right) &= 0 \\
\Gamma^t_{xz} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{zt}}{\partial x} + \frac{\partial g_{tx}}{\partial z} - \frac{\partial g_{xz}}{\partial t}\right) &= 0 \\
\Gamma^t_{yy} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{yt}}{\partial y} + \frac{\partial g_{ty}}{\partial y} - \frac{\partial g_{yy}}{\partial t}\right) &= 0 \\
\Gamma^t_{yz} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{jt}}{\partial y} + \frac{\partial g_{ty}}{\partial z} - \frac{\partial g_{yz}}{\partial t}\right) &= 0 \\
\Gamma^t_{zz} &= \frac{1}{2}g^{tt}\left(\frac{\partial g_{zt}}{\partial z} + \frac{\partial g_{tz}}{\partial z} - \frac{\partial g_{zz}}{\partial t}\right) &= 0
\end{aligned}$$

$$\begin{aligned}
\Gamma^x_{tt} &= \frac{1}{2}g^{xx}\left(\frac{\partial g_{tx}}{\partial t} + \frac{\partial g_{xt}}{\partial t} - \frac{\partial g_{tt}}{\partial x}\right) &= -\frac{1}{2}g^{xx}\frac{\partial g_{tt}}{\partial x} \\
\Gamma^x_{tx} &= \frac{1}{2}g^{xx}\left(\frac{\partial g_{xx}}{\partial t} + \frac{\partial g_{xt}}{\partial x} - \frac{\partial g_{tx}}{\partial x}\right) &= 0 \\
\Gamma^x_{ty} &= \frac{1}{2}g^{xx}\left(\frac{\partial g_{yx}}{\partial t} + \frac{\partial g_{xt}}{\partial y} - \frac{\partial g_{ty}}{\partial x}\right) &= 0 \\
\Gamma^x_{tz} &= \frac{1}{2}g^{xx}\left(\frac{\partial g_{zx}}{\partial t} + \frac{\partial g_{xt}}{\partial z} - \frac{\partial g_{tz}}{\partial x}\right) &= 0 \\
\Gamma^x_{xx} &= \frac{1}{2}g^{xx}\left(\frac{\partial g_{xx}}{\partial x} + \frac{\partial g_{xx}}{\partial x} - \frac{\partial g_{xx}}{\partial x}\right) &= \frac{1}{2}g^{xx}\frac{\partial g_{xx}}{\partial x} \\
\Gamma^x_{xy} &= \frac{1}{2}g^{xx}\left(\frac{\partial g_{yx}}{\partial x} + \frac{\partial g_{xx}}{\partial y} - \frac{\partial g_{xy}}{\partial x}\right) &= \frac{1}{2}g^{xx}\frac{\partial g_{xx}}{\partial y} \\
\Gamma^x_{xz} &= \frac{1}{2}g^{xx}\left(\frac{\partial g_{zx}}{\partial x} + \frac{\partial g_{xx}}{\partial z} - \frac{\partial g_{xz}}{\partial x}\right) &= \frac{1}{2}g^{xx}\frac{\partial g_{xx}}{\partial z} \\
\Gamma^x_{yy} &= \frac{1}{2}g^{xx}\left(\frac{\partial g_{yx}}{\partial y} + \frac{\partial g_{xy}}{\partial y} - \frac{\partial g_{yy}}{\partial x}\right) &= -\frac{1}{2}g^{xx}\frac{\partial g_{yy}}{\partial x}
\end{aligned}$$

$$
\Gamma^x_{yz} = \frac{1}{2} g^{xx} \left(\frac{\partial g_{zx}}{\partial y} + \frac{\partial g_{xy}}{\partial z} - \frac{\partial g_{yz}}{\partial x} \right) = 0
$$
$$
\Gamma^x_{zz} = \frac{1}{2} g^{xx} \left(\frac{\partial g_{zx}}{\partial z} + \frac{\partial g_{xz}}{\partial z} - \frac{\partial g_{zz}}{\partial x} \right) = -\frac{1}{2} g^{xx} \frac{\partial g_{zz}}{\partial x}
$$

$$
\Gamma^y_{tt} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{ty}}{\partial t} + \frac{\partial g_{yt}}{\partial t} - \frac{\partial g_{tt}}{\partial y} \right) = -\frac{1}{2} g^{yy} \frac{\partial g_{tt}}{\partial y}
$$
$$
\Gamma^y_{tx} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{xy}}{\partial t} + \frac{\partial g_{yt}}{\partial x} - \frac{\partial g_{tx}}{\partial y} \right) = 0
$$
$$
\Gamma^y_{ty} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{yy}}{\partial t} + \frac{\partial g_{yt}}{\partial y} - \frac{\partial g_{ty}}{\partial y} \right) = 0
$$
$$
\Gamma^y_{tz} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{zy}}{\partial t} + \frac{\partial g_{yt}}{\partial z} - \frac{\partial g_{tz}}{\partial y} \right) = 0
$$
$$
\Gamma^y_{xx} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{xy}}{\partial x} + \frac{\partial g_{yx}}{\partial x} - \frac{\partial g_{xx}}{\partial y} \right) = -\frac{1}{2} g^{yy} \frac{\partial g_{xx}}{\partial y}
$$
$$
\Gamma^y_{xy} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{yy}}{\partial x} + \frac{\partial g_{yx}}{\partial y} - \frac{\partial g_{xy}}{\partial y} \right) = \frac{1}{2} g^{yy} \frac{\partial g_{yy}}{\partial x}
$$
$$
\Gamma^y_{xz} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{zy}}{\partial x} + \frac{\partial g_{yx}}{\partial z} - \frac{\partial g_{xz}}{\partial y} \right) = 0
$$
$$
\Gamma^y_{yy} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{yy}}{\partial y} + \frac{\partial g_{yy}}{\partial y} - \frac{\partial g_{yy}}{\partial y} \right) = \frac{1}{2} g^{yy} \frac{\partial g_{yy}}{\partial y}
$$
$$
\Gamma^y_{yz} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{zy}}{\partial y} + \frac{\partial g_{yy}}{\partial z} - \frac{\partial g_{yz}}{\partial y} \right) = \frac{1}{2} g^{yy} \frac{\partial g_{yy}}{\partial z}
$$
$$
\Gamma^y_{zz} = \frac{1}{2} g^{yy} \left(\frac{\partial g_{zy}}{\partial z} + \frac{\partial g_{yz}}{\partial z} - \frac{\partial g_{zz}}{\partial y} \right) = -\frac{1}{2} g^{yy} \frac{\partial g_{zz}}{\partial y}
$$

$$
\Gamma^z_{tt} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{tz}}{\partial t} + \frac{\partial g_{zt}}{\partial t} - \frac{\partial g_{tt}}{\partial z} \right) = -\frac{1}{2} g^{zz} \frac{\partial g_{tt}}{\partial z}
$$
$$
\Gamma^z_{tx} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{xz}}{\partial t} + \frac{\partial g_{zt}}{\partial x} - \frac{\partial g_{tx}}{\partial z} \right) = 0
$$
$$
\Gamma^z_{ty} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{yz}}{\partial t} + \frac{\partial g_{zt}}{\partial y} - \frac{\partial g_{ty}}{\partial z} \right) = 0
$$
$$
\Gamma^z_{tz} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{zz}}{\partial t} + \frac{\partial g_{zt}}{\partial z} - \frac{\partial g_{tz}}{\partial z} \right) = 0
$$
$$
\Gamma^z_{xx} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{xz}}{\partial x} + \frac{\partial g_{zx}}{\partial x} - \frac{\partial g_{xx}}{\partial z} \right) = -\frac{1}{2} g^{zz} \frac{\partial g_{xx}}{\partial z}
$$
$$
\Gamma^z_{xy} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{yz}}{\partial x} + \frac{\partial g_{zx}}{\partial y} - \frac{\partial g_{xy}}{\partial z} \right) = 0
$$
$$
\Gamma^z_{xz} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{zz}}{\partial x} + \frac{\partial g_{zx}}{\partial z} - \frac{\partial g_{xz}}{\partial z} \right) = \frac{1}{2} g^{zz} \frac{\partial g_{zz}}{\partial x}
$$
$$
\Gamma^z_{yy} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{yz}}{\partial y} + \frac{\partial g_{zy}}{\partial y} - \frac{\partial g_{yy}}{\partial z} \right) = -\frac{1}{2} g^{zz} \frac{\partial g_{yy}}{\partial z}
$$
$$
\Gamma^z_{yz} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{zz}}{\partial y} + \frac{\partial g_{zy}}{\partial z} - \frac{\partial g_{yz}}{\partial z} \right) = \frac{1}{2} g^{zz} \frac{\partial g_{zz}}{\partial y}
$$
$$
\Gamma^z_{zz} = \frac{1}{2} g^{zz} \left(\frac{\partial g_{zz}}{\partial z} + \frac{\partial g_{zz}}{\partial z} - \frac{\partial g_{zz}}{\partial z} \right) = \frac{1}{2} g^{zz} \frac{\partial g_{zz}}{\partial z}
$$

Note that there is no magic or hidden step in how we compute the Christoffel symbols, computing the Christoffel symbols just becomes a game of spotting all off-diagonal elements, i.e. g_{ij} where $i \neq j$, and removing them since they are zero and keeping all the diagonal elements. Additionally the derivative of the metric with respect to the first coordinate (time, t) is zero so those elements also disappear. The Christoffel symbols are thus entirely written in known functions.

If we write out the matrices of the Christoffel symbols we can see a clear structure

$$
\Gamma^t = \frac{g^{tt}}{2} \begin{bmatrix} & \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_1}{\partial x} & & & \\ \frac{\partial f_1}{\partial y} & & & \\ \frac{\partial f_1}{\partial z} & & & \end{bmatrix}
$$

$$
\Gamma^x = \frac{g^{xx}}{2} \begin{bmatrix} \frac{\partial f_1}{\partial x} & & & \\ & \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ & \frac{\partial f_2}{\partial y} & -\frac{\partial f_2}{\partial x} & \\ & \frac{\partial f_2}{\partial z} & & -\frac{\partial f_2}{\partial x} \end{bmatrix}
$$

$$
\Gamma^y = \frac{g^{yy}}{2} \begin{bmatrix} \frac{\partial f_1}{\partial y} & & & \\ & -\frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial x} & \\ & \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ & & \frac{\partial f_2}{\partial z} & -\frac{\partial f_2}{\partial y} \end{bmatrix}
$$

$$
\Gamma^z = \frac{g^{zz}}{2} \begin{bmatrix} \frac{\partial f_1}{\partial z} & & & \\ & -\frac{\partial f_2}{\partial z} & & \frac{\partial f_2}{\partial x} \\ & & -\frac{\partial f_2}{\partial z} & \frac{\partial f_2}{\partial z} \\ & \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \end{bmatrix}
$$

With the Christoffel symbols computed we can rewrite the geodesic equation in the form required for the numerical integration

$$
\ddot{\gamma}^k = -\dot{\gamma}^{\mathrm{T}} \Gamma^k \dot{\gamma}.
$$

This can now be implemented with some appropriate Runge-Kutta method to approximate geodesics on the space-time manifold.

Choosing a Runge-Kutta Method

The de facto standard adaptive Runge-Kutta method for approximating solutions to differential equations is the Runge-Kutta45 method which is a fifth order method with an embedded fourth order method to approximate the error. For our purposes this might let us take far too large steps which we can see below.

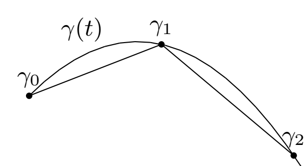

Figure 3: Above is pictured a true geodesic, $\gamma(t)$, and the approximation by a discrete set of points γ_0, γ_1 and so on.

In figure 3 we see an illustration of why taking too large steps might be undesirable, when we compute if a geodesic collides with an object in the scene we can only compute the collision point with respect to the linear interpolation between two discrete points rather than the true path. For this reason any collision point we compute will necessarily be incorrect and the error will be larger the larger the step size is.

For this reason we will step down one order of convergence and use a fourth order convergent Runge-Kutta method with an embedded third order convergent method which can be seen in the following table:

$$
\begin{array}{c|ccccc}
0 & & & & & \\
0.5 & 0.5 & & & & \\
0.5 & & 0.5 & & & \\
0.75 & & 0.75 & & & \\
1 & & & 1 & & \\
\hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{6} \\
 & \frac{2}{9} & \frac{1}{3} & 0 & \frac{4}{9} & 0
\end{array}
$$

These methods are not random and are quite standard although this particular combination to produce an embedded method might be non-standard while still being perfectly valid.

Once we have decided on a Runge-Kutta method we can use the fact that we are solving a second order differential equation to precompute parts of the update rule. Let the initial position and tangent of the geodesic be

$$
\begin{aligned}
\gamma_0 &= p \\
\dot{\gamma}_0 &= v
\end{aligned}
$$

we then define the following function and variables

$$
\begin{aligned}
f(\dot{\gamma}) &= \begin{bmatrix} -\dot{\gamma}^{\mathrm{T}}\Gamma^{t}\dot{\gamma} \\ -\dot{\gamma}^{\mathrm{T}}\Gamma^{x}\dot{\gamma} \\ -\dot{\gamma}^{\mathrm{T}}\Gamma^{y}\dot{\gamma} \\ -\dot{\gamma}^{\mathrm{T}}\Gamma^{z}\dot{\gamma} \end{bmatrix} \\
k_1 &= f(\gamma_i) \\
k_2 &= f\left(\gamma_i + \frac{h}{2}k_1\right) \\
k_3 &= f\left(\gamma_i + \frac{h}{2}k_2\right) \\
k_3^* &= f\left(\gamma_i + \frac{3h}{4}k_2\right) \\
k_4 &= f\left(\gamma_i + hk_3\right)
\end{aligned}
$$

the update rule can then be written in the following way

$$
\begin{bmatrix} \gamma_{i+1} \\ \dot{\gamma}_{i+1} \end{bmatrix} = \begin{bmatrix} \gamma_i \\ \dot{\gamma}_i \end{bmatrix} + \begin{bmatrix} \left(1 + \frac{2}{3}h\right)\dot{\gamma}_i \\ \frac{h}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right) \end{bmatrix}
$$

with the error estimate

$$
e_{i+1} = \begin{bmatrix} \frac{h^2}{6}\dot{\gamma} \\ \frac{h}{6}\left(-k_1 + 2k_3 - 4k_3^* + k_4\right) \end{bmatrix}.
$$

Ray-Scene Intersection

Once we have a way to approximate geodesics we can, from a point and a direction, compute a new point on the geodesic To then compute an intersection point we interpolate linearly between the two most recent points. In figure 4 we can see how, even if the points we compute lie exactly on the geodesic we are approximating a large step size might lead to large errors in the determined intersection points.

For this reason we have to choose an appropriate error tolerance which does not give too large of a step size such that the intersection points are substantially incorrect but not too small of a step size such that each ray takes too long to compute.

The chosen error tolerances are the following

$$
\begin{aligned}
\tau &= 0.01 \\
\tau_1 &= 0.001 \\
\tau_2 &= 0.1
\end{aligned}
$$

to make sure that the adaptive method actually stays within these error tolerances it was compared to a Runge-Kutta method with a fixed very small step size and the error was actually several orders of magnitude lower than the desired error τ.

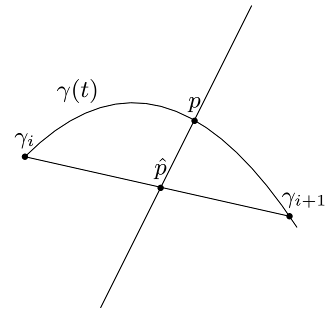

Figure 4: In the figure we see a geodesic, $\gamma(t)$, two points that lie on the geodesic and a linear interpolation between them. The intersection points with some object has been marked. The true intersection point is p and the approximate intersection point is $\hat{p}$.

Results

The code which implements a relativistic ray marcher can be found at the following link github.com/p-rosit/Relativistic_Ray_Marcher and some examples can be seen in figure 5, 6, 7, 8 and 9. In figure 5 we can see the black hole with a flat disk around it, the disk can be seen above the black hole due to the fact that light rays that go above the black hole are bent downwards and hit the disk from above.

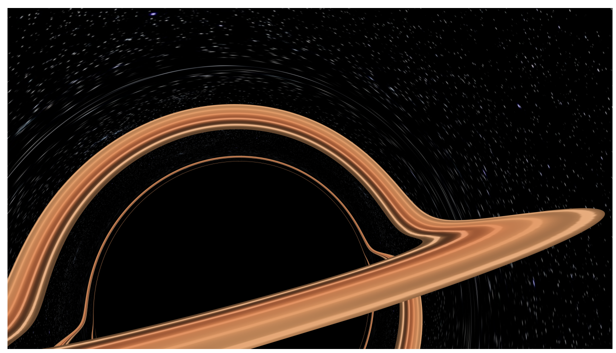

Figure 5: Black hole with a disk around it seen from up close.

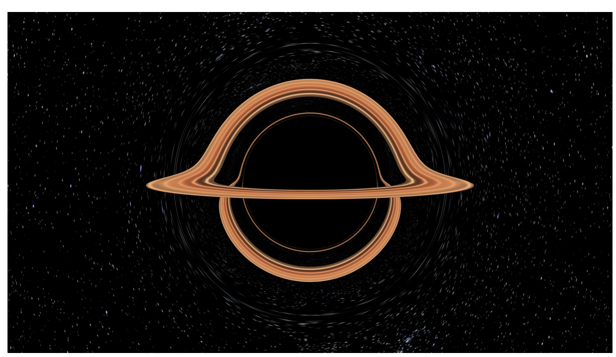

Figure 6: Black hole with disk around it seen from the "front".

In figure 8 we again see the black hole but this time the camera is located 3 units of distance from the black hole and points towards the origin, where the black hole is located. If we just look at these two images it is hard to get a sense of what the scene actually looks like and what objects are present in the scene since humans use the image of a scene to build up a 3d model of what the scene looks like. To compare we have also rendered the scene where the light does not bend which can be seen in figure 7.

In this image we see that the event horizon is, in reality quite small in the image but since light rays are drawn towards it light rays that would otherwise not have passed the event horizon hit it anyway. The disk can also be seen and it is, clearly, flat.

In figure 8 we note thus that we are actually seeing several disks. This is due to the fact that light revolves around the black hole one or more times. This effect can most clearly be seen in figure 5 where we can see the disk several times. If we were to zoom in infinitely far we would be able to see an infinte amount of copies of the disk and we alternate between seeing it from above and below.

Figure 7: Black hole seen from the "front" but light does not bend.

In figure 8 and 9 the same scene has been rendered with different textures where it becomes evident that the event horizon of the black hole is a sphere and the scene is enclosed in a cube.

Figure 8: Black hole seen from the "front" with textures that make the bending of light more apparent.

The grey lines present on the disk are uniformly distributed and from figure 8 we can see that above the black hole the disk has been stretched out which means that a relatively small part of the disk constitutes the top part of the disk.

We can also see that both poles of the sphere can be seen in image 8 neither of which can be seen in image 9 and in particular the entire surface of the sphere can actually be seen in image 8. Again, if we zoom in to the border of the sphere in image 8 we will see the entire surface of the event horizon an infinite amount of times.

In figure 9 we can see that the blue wall has a texture of a gradient that is light blue on the bottom and dark blue on the top. With this knowledge we can in figure 8 see that just above the disk the wall is light blue which means that those light rays have hit the bottom of the wall, this strengthens our claim that the part of the disk that is seen above the black hole is the top of the disk behind the black hole. Similarly we see that below the disk below the black hole the

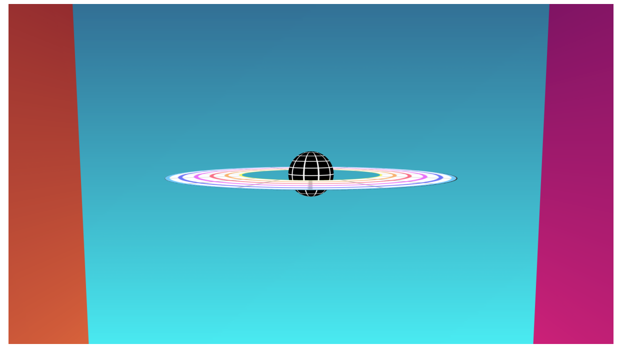

Figure 9: Black hole seen from the "front" with where light does not bend with the same textures as figure 8.

wall is dark blue which means that those light rays are hitting the top of the wall which also strengthens our claim that we are seeing the alternating copies of the top and bottom of the disk.

The same adaptive Runge-Kutta method which solved the geodesic equation was set up to solve the Jacobi equation but this turned out to not be tractable. The Jacobi equation is, or at least seems to be, a stiff differential equation which would mean that an explicit Runge-Kutta method is not appropriate to solve the Jacobi equation. A stiff differential equation has no proper definition but are characterized by the fact that unreasonably small step sizes are needed to satisfy error tolerances as the solution is approximated.

Discussion

The renderer is able to render disks, spheres and planes because those are very easy to texture and their intersection with a line segment is very easy to compute. In reality we would only actually have to implement ray-triangle intersection to be able to image this scene but this would have led to difficulties in implementation of functions that were not the focus of this project.

We could, however, quite easily implement ray-triangle intersection which would enable us to image more general scenes but for each step of the ray we have to check if it collides with anything, by keeping track of which objects were close to hitting the ray last time as well as collecting the objects in a bounding volume hierarchy this renderer could be greatly sped up if many objects are to be rendered.

This renderer is physically correct as long as the camera is small due to the fact that rays are not spawned from the camera centre, they are rather spawned from the position of the pixel in world space they are associated with. To be strictly physically correct we would have to approximate the geodesic between the camera position and the position on the camera plane to initialise the rays correctly.

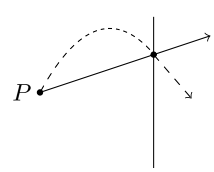

Figure 10: A light ray shot from the camera, P, where the light does not bend can be seen in the full line and its tangent has been drawn at that point. A light ray which bends has also been shot and is the dashed line and its tangent is the dashed arrow. Both of these rays intersect the camera plane at the same point.

In figure 10 we can see why it is important to compute the geodesic from the where the ray is shot to where it intersects the camera plane. If we do not compute the correct geodesic we will not get the correct tangent for the ray we shoot from the camera and the image will be wrong, this will only affect the image when the camera is close to the black hole or if the camera is big.

The problem with finding the geodesic between two points is that this is a boundary value problem rather than an inital value problem, these are much harder to solve efficiently but it is still possible by solving a non-linear system of equations with Newton's method or similar.

Future Work

In future work it would be interesting to implement this ray marcher to work on a GPU as well as to fix the problem of finding the correct geodesic between the camera position and a desired point on the camera plane.

It would also be interesting to implement a bounding volume hierarchy and ray-triangle intersections to be able to image more general scenes.

At the moment the renderer only supports diffuse surfaces but it would be interesting to implement reflections and refractions, light sources as well as glow since the disk around the black hole would look better if it glowed.

The metric induced by a rotating black hole has also been derived and it is called the Kerr metric and light bends quite differently for these types of black hole. The metric has a far less friendly structure and deriving the Christoffel symbols becomes far more tedious but it would be very interesting to implement a ray marcher which images scenes induces by this metric.

In general the metric on the space-time manifold comes from the Einstein field equations which is a set of partial differential equations which curve the manifold depending on many different parameters. If would be interesting to approximate the metric with some finite element method and be able to not just image a scene with a singular black hole but also a scene with several black holes that are perhaps rotating as well as rotating around each other. This is however most likely a project worthy of a PHD and will not be happening any time soon.

References

[1] Sigmundur Gudmundsson. *An Introduction to Riemannian Geometry*. Downloaded in December 2022. 2022. URL: https://www.matematik.lu.se/matematiklu/personal/sigma/Riemann.pdf.

[2] Arieh Iserles. *A First Course in Numerical Analysis of Differential Equations*. 2014.

Appendix

The rewriting of the Jacobi equation to a form more friendly for numerical computations is quite tedious but for completeness we include it here. Remember that the Jacobi equation is the following

$$\nabla_{\dot\gamma(t)}\nabla_{\dot\gamma(t)}J(\gamma(t)) + R(J(\gamma(t)),\dot\gamma(t))\dot\gamma(t) = 0$$

note that the Riemannian curvature tensor R has the following coordinates

$$R^k_{lij} = \frac{\partial}{\partial i}\Gamma^k_{jl} - \frac{\partial}{\partial j}\Gamma^k_{il} + \sum_{m=1}^{n}\Gamma^k_{im}\Gamma^m_{jl} - \Gamma^k_{jm}\Gamma^m_{il}$$

computing these functions if extremely tedious but to check if computing them by hand was worth it these functions were first computed via numerical differentiation which is why the functions for the Riemannian curvature tensor is not included in this report.

The second term of the Jacobi equation can be written in the following way

$$\begin{aligned} R(J(\gamma(t)),\dot\gamma(t))\dot\gamma(t) &= \sum_{m=1}^{n}\sum_{l,i,j=1}^{n} R^k_{lij}J^i\dot\gamma^j\dot\gamma^l\frac{\partial}{\partial x_m} \\ &= \sum_{m=1}^{n}\sum_{l=1}^{n}\dot\gamma^l\left(J^{\mathsf{T}}R^k_l\dot\gamma\right)\frac{\partial}{\partial x_m} \end{aligned}$$

where we have defined the matrix R^k_l such that the element at position i,j is R^k_{lij}.

Now consider the the following expression:

$$\nabla_{\dot\gamma}J = \sum_k\left(\dot J_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\frac{\partial}{\partial x_k}$$

which we know is correct since the derivation is the same as how we derived the expression for $\nabla_{\dot\gamma}\dot\gamma$. Then we can calculate the iterated covariant derivative of the Jacobi field J:

$$\begin{aligned}
\nabla_{\dot\gamma}\nabla_{\dot\gamma}J &= \nabla_{\dot\gamma}\left(\sum_k\left(\dot J_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\frac{\partial}{\partial x_k}\right) \\
&= \sum_{k=1}^{n}\left(\dot\gamma\left(\dot J_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\frac{\partial}{\partial x_k} + \right. \\
&\qquad \left. + \left(\dot J_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\nabla_{\dot\gamma}\frac{\partial}{\partial x_k}\right) \\
&= \sum_{k=1}^{n}\left(\dot\gamma\left(\dot J_k\right)\frac{\partial}{\partial x_k} + \sum_{i,j}\dot\gamma\left(J_iJ_j\Gamma^k_{ij}\right)\frac{\partial}{\partial x_k}\right. \\
&\qquad \left. + \left(\dot J_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\left(\sum_{a=1}^{n}\dot\gamma_a\nabla_{\frac{\partial}{\partial x_a}}\frac{\partial}{\partial x_k}\right)\right) \\
&= \sum_{k=1}^{n}\left(\ddot J_k\frac{\partial}{\partial x_k} + \sum_{i,j}\dot\gamma\left(J_iJ_j\Gamma^k_{ij}\right)\frac{\partial}{\partial x_k}\right. \\
&\qquad \left. + \left(\dot J_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\left(\sum_{a=1}^{n}\dot\gamma_a\sum_{b=1}^{n}\Gamma^b_{ak}\frac{\partial}{\partial x_b}\right)\right) \\
&= \sum_{k=1}^{n}\left(\left(\ddot J_k + \sum_{i,j}\dot\gamma\left(J_iJ_j\Gamma^k_{ij}\right)\right)\frac{\partial}{\partial x_k}\right. \\
&\qquad \left. + \left(\dot J_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\left(\sum_{a,b}\dot\gamma_a\Gamma^b_{ak}\frac{\partial}{\partial x_b}\right)\right) \\
&= \sum_{b=1}^{n}\left(\left(\ddot J_b + \sum_{i,j}\dot\gamma\left(J_iJ_j\Gamma^b_{ij}\right)\right)\frac{\partial}{\partial x_b}\right. \\
&\qquad \left. + \sum_{a,k}\left(\dot J_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\left(\dot\gamma_a\Gamma^b_{ak}\frac{\partial}{\partial x_b}\right)\right) \\
&= \sum_{b=1}^{n}\left(\ddot J_b + \sum_{i,j}\dot\gamma\left(J_iJ_j\Gamma^b_{ij}\right)\right. \\
&\qquad \left. + \sum_{a,k}\left(\dot J_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\left(\dot\gamma_a\Gamma^b_{ak}\right)\right)\frac{\partial}{\partial x_b}
\end{aligned}$$

Here we need to compute the result of the vector field $\dot\gamma$ applied to the function $J_iJ_j\Gamma^b_{ij}$ which is the directional derivative along the tangent. To compute this we need to apply the product rule:

$$\begin{aligned}\dot{\gamma}\left(J_iJ_j\Gamma^b_{ij}\right) &= \dot{\gamma}\left(J_i\right)J_j\Gamma^b_{ij} + J_i\dot{\gamma}\left(J_j\right)\Gamma^b_{ij} + J_iJ_j\dot{\gamma}\left(\Gamma^b_{ij}\right) \\ &= \dot{J}_iJ_j\Gamma^b_{ij} + J_i\dot{J}_j\Gamma^b_{ij} + J_iJ_j\dot{\gamma}\left(\Gamma^b_{ij}\right)\end{aligned}$$

here we also need to compute the result of $\dot{\gamma}\left(\Gamma^b_{ij}\right)$ which we can do in the following way

$$\begin{aligned}\dot{\gamma}\left(\Gamma^b_{ij}\right) &= \left(\sum_{m=1}^{n}\dot{\gamma}^m\frac{\partial}{\partial x_m}\right)\left(\Gamma^b_{ij}\right) \\ &= \sum_{m=1}^{n}\dot{\gamma}^m\frac{\partial\Gamma^b_{ij}}{\partial x_m}\end{aligned}$$

Thus we see that the full formula becomes

$$\begin{aligned}\nabla_{\dot{\gamma}}\nabla_{\dot{\gamma}}J = \sum_{b=1}^{n}\Bigg(\ddot{J}_b + \sum_{i,j}\left(\dot{J}_iJ_j\Gamma^b_{ij} + J_i\dot{J}_j\Gamma^b_{ij} + J_iJ_j\sum_{m=1}^{n}\dot{\gamma}^m\frac{\partial\Gamma^b_{ij}}{\partial x_m}\right) \\ + \sum_{a,k}\left(\dot{J}_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\left(\dot{\gamma}_a\Gamma^b_{ak}\right)\Bigg)\frac{\partial}{\partial x_b}\end{aligned}$$

The Jacobi equation is thus the following

$$\begin{aligned}0 = \sum_{b=1}^{n}\Bigg(\ddot{J}_b + \sum_{i,j}\left(\dot{J}_iJ_j\Gamma^b_{ij} + J_i\dot{J}_j\Gamma^b_{ij} + J_iJ_j\sum_{m=1}^{n}\dot{\gamma}^m\frac{\partial\Gamma^b_{ij}}{\partial x_m}\right) \\ + \sum_{a,k}\left(\dot{J}_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\left(\dot{\gamma}_a\Gamma^b_{ak}\right) + \sum_{l}\dot{\gamma}^l\left(J^{\mathrm{T}}R^b_l\dot{\gamma}\right)\Bigg)\frac{\partial}{\partial x_b}\end{aligned}$$

this formula should be equal to zero which means that every coordinate has to be zero which means that we get the following set of n equations

$$\begin{aligned}0 = \ddot{J}_b + \sum_{i,j}\left(\dot{J}_iJ_j\Gamma^b_{ij} + J_i\dot{J}_j\Gamma^b_{ij} + J_iJ_j\sum_{m=1}^{n}\dot{\gamma}^m\frac{\partial\Gamma^b_{ij}}{\partial x_m}\right) \\ + \sum_{a,k}\left(\dot{J}_k + \sum_{i,j}J_iJ_j\Gamma^k_{ij}\right)\left(\dot{\gamma}_a\Gamma^b_{ak}\right) + \sum_{l}\dot{\gamma}^l\left(J^{\mathrm{T}}R^b_l\dot{\gamma}\right)\end{aligned}$$

many of these terms can be simplified by writing them as matrix products instead since $v^{\mathrm{T}}Av = \sum_{ij}v_iA_{ij}v_j$, thus we can write

$$\begin{aligned}0 &= \ddot{J}_b + \dot{J}^{\mathrm{T}}\Gamma^bJ + J^{\mathrm{T}}\Gamma^b\dot{J} + J^{\mathrm{T}}\left(\sum_m\dot{\gamma}^m\frac{\partial}{\partial x_m}\Gamma^b\right)J \\ &\quad + \sum_{a,k}\left(\dot{J}_k + J^{\mathrm{T}}\Gamma^kJ\right)\left(\dot{\gamma}_a\Gamma^b_{ak}\right) + \sum_l\dot{\gamma}^l\left(J^{\mathrm{T}}R^b_l\dot{\gamma}\right) \\ &= \ddot{J}_b + 2\dot{J}^{\mathrm{T}}\Gamma^bJ + J^{\mathrm{T}}\left(\sum_m\dot{\gamma}^m\frac{\partial}{\partial x_m}\Gamma^b\right)J \\ &\quad + \sum_{a,k}\left(\dot{J}_k\dot{\gamma}_a\Gamma^b_{ak} + J^{\mathrm{T}}\Gamma^kJ\dot{\gamma}_a\Gamma^b_{ak}\right) + \sum_l\dot{\gamma}^l\left(J^{\mathrm{T}}R^b_l\dot{\gamma}\right) \\ &= \ddot{J}_b + 2\dot{J}^{\mathrm{T}}\Gamma^bJ + J^{\mathrm{T}}\left(\sum_m\dot{\gamma}^m\frac{\partial}{\partial x_m}\Gamma^b\right)J \\ &\quad + \dot{J}^{\mathrm{T}}\Gamma^b\dot{\gamma} + \left(\sum_k\left(J^{\mathrm{T}}\Gamma^kJ\right)\left(\Gamma^b\dot{\gamma}\right)\right) + \sum_l\dot{\gamma}^l\left(J^{\mathrm{T}}R^b_l\dot{\gamma}\right) \\ &= \ddot{J}_b + 2\dot{J}^{\mathrm{T}}\Gamma^bJ + J^{\mathrm{T}}\left(\sum_m\dot{\gamma}^m\frac{\partial}{\partial x_m}\Gamma^b\right)J \\ &\quad + \dot{J}^{\mathrm{T}}\Gamma^b\dot{\gamma} + \left(\Gamma^b\dot{\gamma}\right)\sum_k\left(J^{\mathrm{T}}\Gamma^kJ\right) + \sum_l\dot{\gamma}^l\left(J^{\mathrm{T}}R^b_l\dot{\gamma}\right)\end{aligned}$$

Finally we see that the complete formula for the Jacobi equation is the following:

$$\begin{aligned}0 = \ddot{J}_k + 2\dot{J}^{\mathrm{T}}\Gamma^kJ + J^{\mathrm{T}}\left(\sum_m\dot{\gamma}^m\frac{\partial}{\partial x_m}\Gamma^k\right)J \\ + \dot{J}^{\mathrm{T}}\Gamma^k\dot{\gamma} + \left(\Gamma^k\dot{\gamma}\right)\sum_m\left(J^{\mathrm{T}}\Gamma^mJ\right) + \sum_l\dot{\gamma}^l\left(J^{\mathrm{T}}R^k_l\dot{\gamma}\right)\end{aligned}$$

due to the laborious nature of deriving this formula and the fact that I have not been able to find it in any literature it would not be too surprising if some logical error has been able to sneak into this formula but it is, to my knowledge, correct.