①

For this model gradient descent was used due to the limitations of calculating the Hessian and therefor a loss function was implemented.

loss function = datapoints − P(A12, A21)

$$\min_{A21, A12} \| \text{loss function} \|^2$$

Method

While gradient < 0.01 :

~~line search A12, A21~~

~~for~~

get direction that will reduce loss value the most.
take a ~~tp~~ step in that direction
calculate new A12, A21
calculate new gradient.
if criteria is not met repeat.

Solution: graph "curve_fitting.png"

the initial guess was A12 = 1, A21 = 1
once iterations were completed ~~and~~ it converged at A12 = 1.952, A21 = 1.694

When looking at the curve fit it can be seen that it fits quite well.