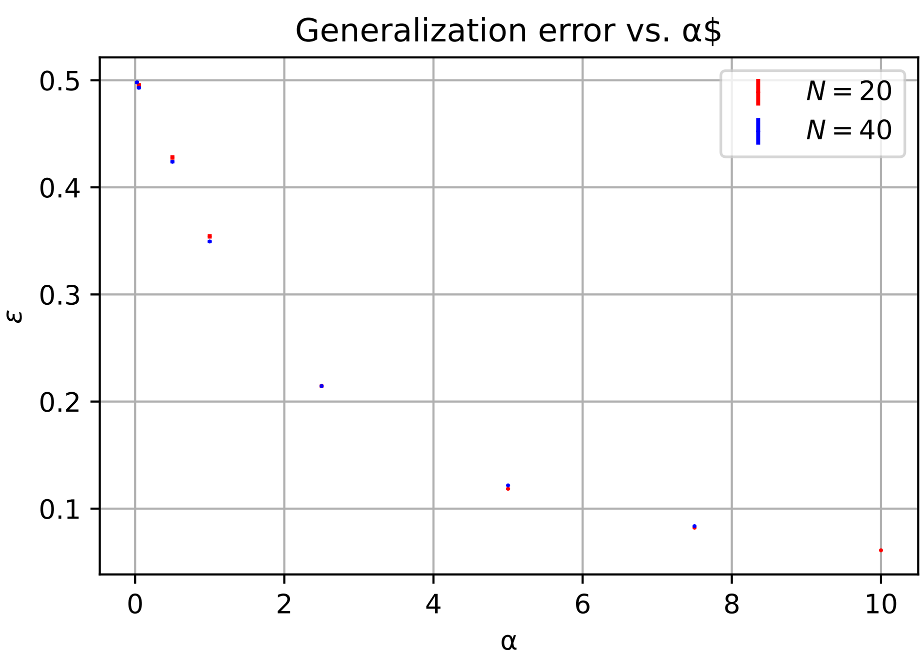
Exercise 04: The annealed solution to randomised perceptron learning

We now take the numerical results of Exercise 3 and compare them to the analytical behaviour that we would expect. In particular, we are studying the behaviour of the generalization error ϵ as $\alpha = P/N$ grows.

The analytic expression is the following

$$\alpha = \frac{(1-\epsilon)\pi}{\tan(\pi\epsilon)}$$

The data given by the previous exercise are plotted here.



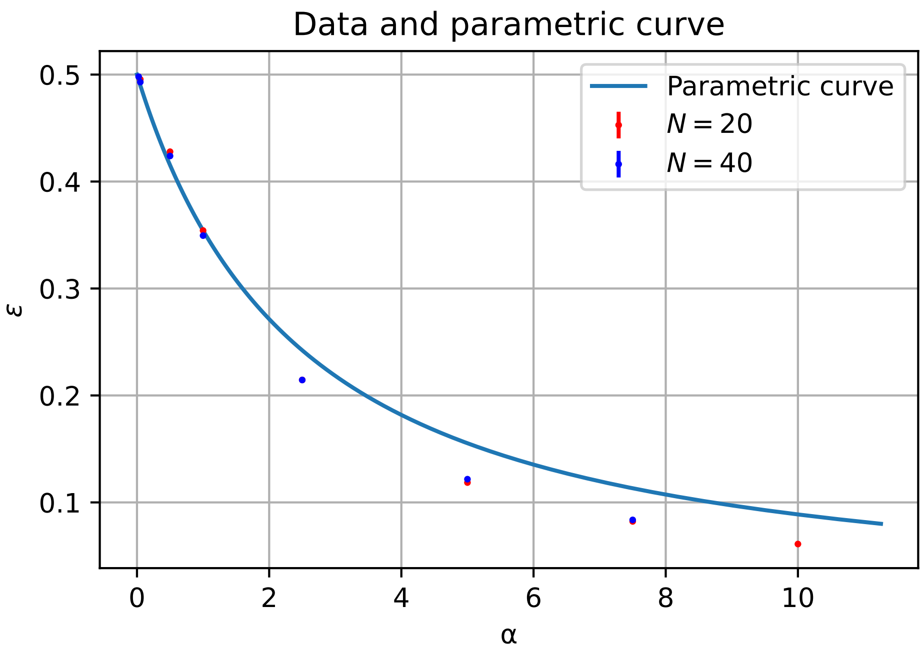
Step 1: Plot the parametric curve

We want parametrize $\alpha = \alpha(t)$ and $\epsilon = \epsilon(t)$ with parameter $t \in [0, 0.5]$.

Let's take

$$\alpha(t) = \frac{(1-t)\pi}{\tan(\pi t)}$$

and $\epsilon(t) = t$.



Step 2: Consider a strategy to solve iteratively the eqn for fixed α

Let us devise an iterative algorithm in order to find a solution for a certain α.

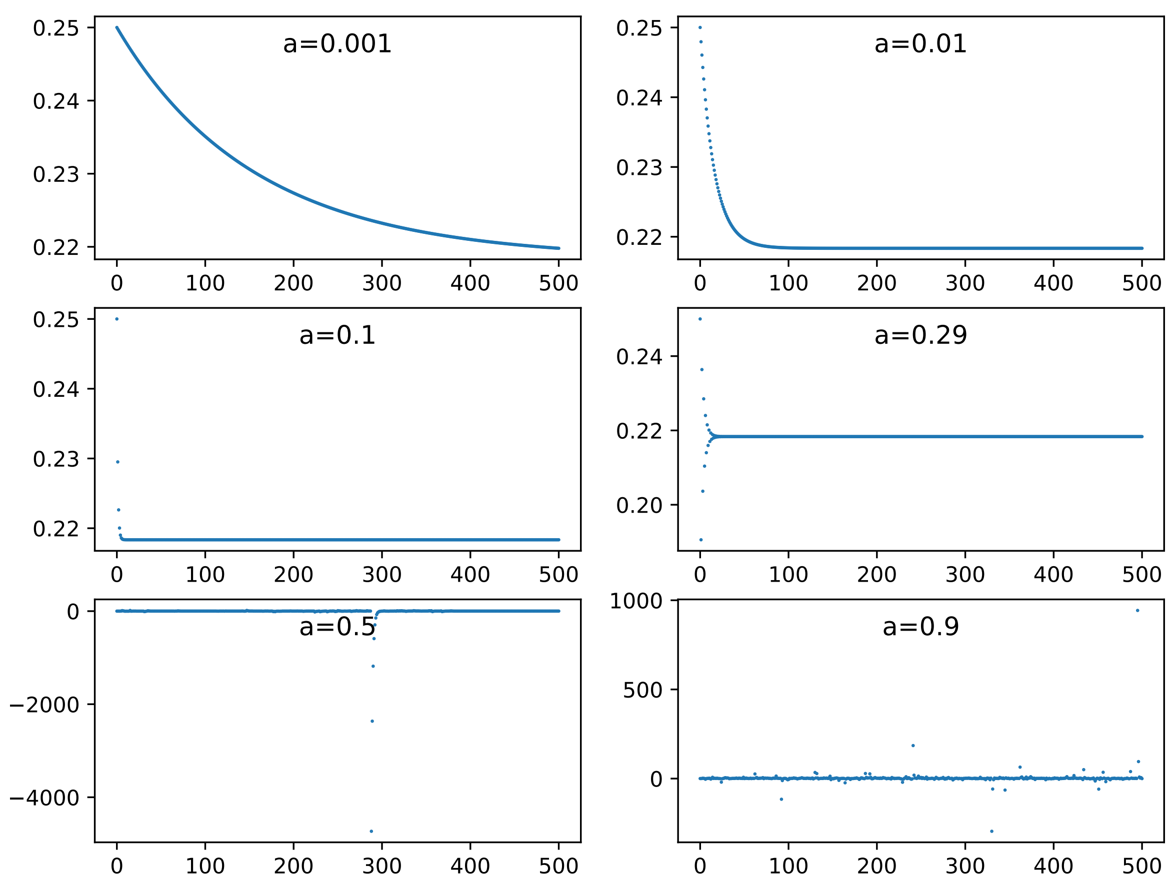
Say, for example, that $\alpha = 5$. Given a self-consistent function of the form $\epsilon = f(\epsilon, a)$, we can:

- Set an initial value $\epsilon_0 = 0.25$, for example, which is a good value in the codomain that we expect.
- Evaluate iteratively $\epsilon_{i+1} = (1-a)\epsilon_i + af(\epsilon_i, a)$
- Plot how ϵ_i changes with i to check for convergence at different settings of a.

Here we have chosen, for example:

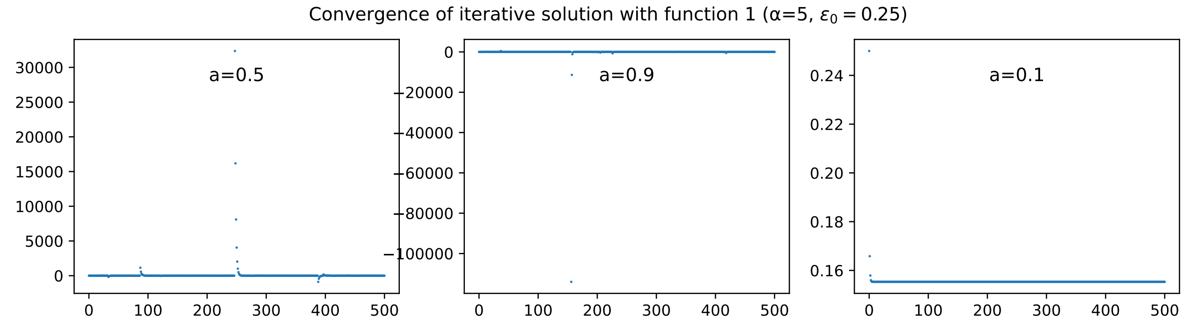
$$f_2(\epsilon, \alpha) = 1 - \frac{\alpha \tan(\pi\epsilon)}{\pi}$$

Convergence of iterative solution (α=5, $\varepsilon_0 = 0.25$)



Step 3: Check that a possible choice is $f = 1 - \alpha \tan(\pi\epsilon)/\pi$

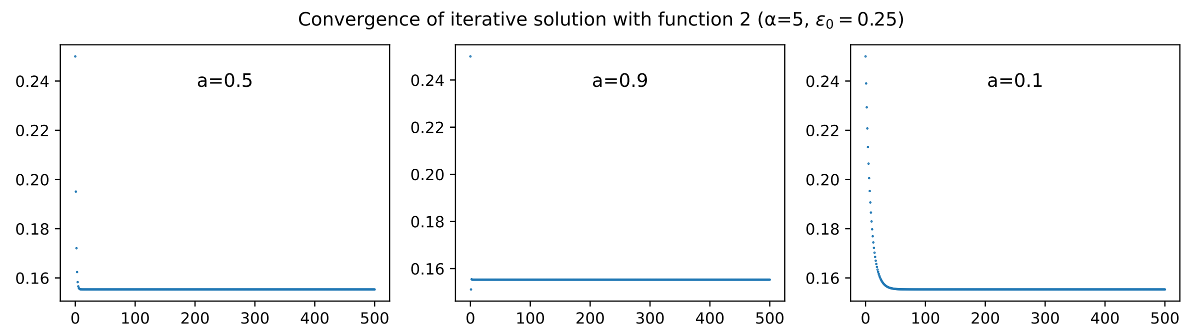
This is similar to the previous case, apart from the choice of α. Let's check for the specific plots we are asked for.



Step 4: Check that a possible choice is $f = \arctan(\pi(1-\epsilon)/\alpha)/\pi$

Let's see what happens with the following choice of the function, and the same parameters.

$$f_1(\epsilon, \alpha) = \frac{1}{\pi}\arctan(\frac{\pi(1-\epsilon)}{\alpha})$$



Step 5 and 6: How do you explain the results?

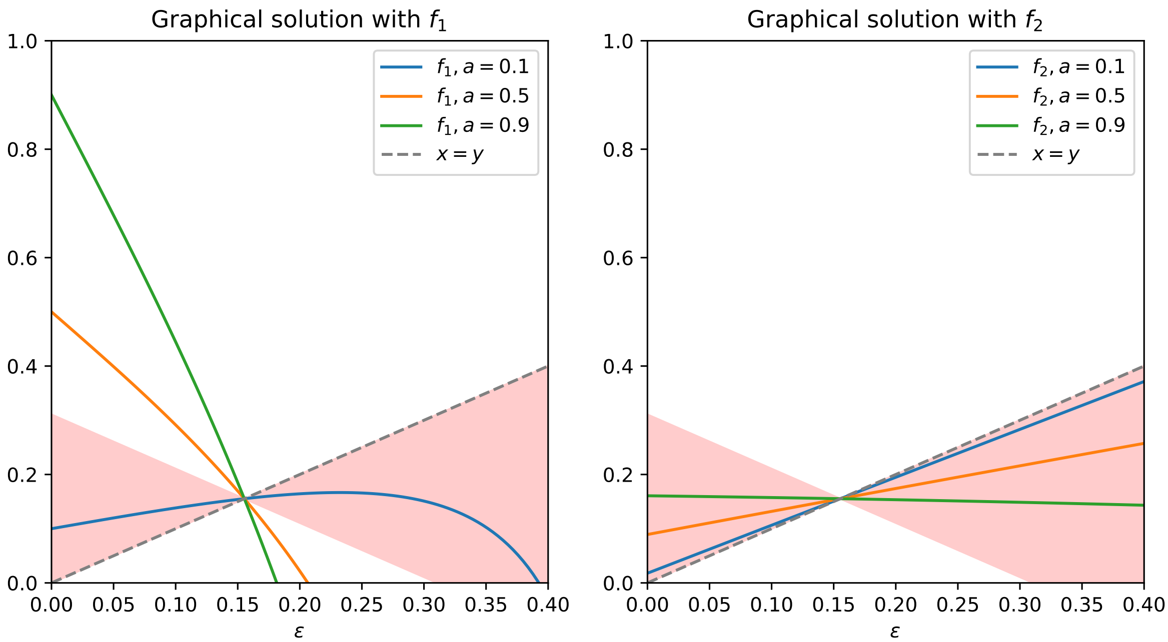
We see that in both cases the iterative algorithm reaches convergence, but not for all values of a. For f_1, for example, we reach convergence only for the lowest value ($a = 1$). Meanwhile in the f_2 case we reach convergence in all the three cases. So **we deem f_2 to be a better candidate when looking for the solutions of the equation**.

We know from the **Banach theorem** that we have guaranteed convergence over a recursive function $x = G(x)$ towards a fixed point if the fixed point x^* exists, and if the absolute value of the derivative of the function is such that $|\frac{dG(x)}{dx}|_{x=x^*} < 1$.

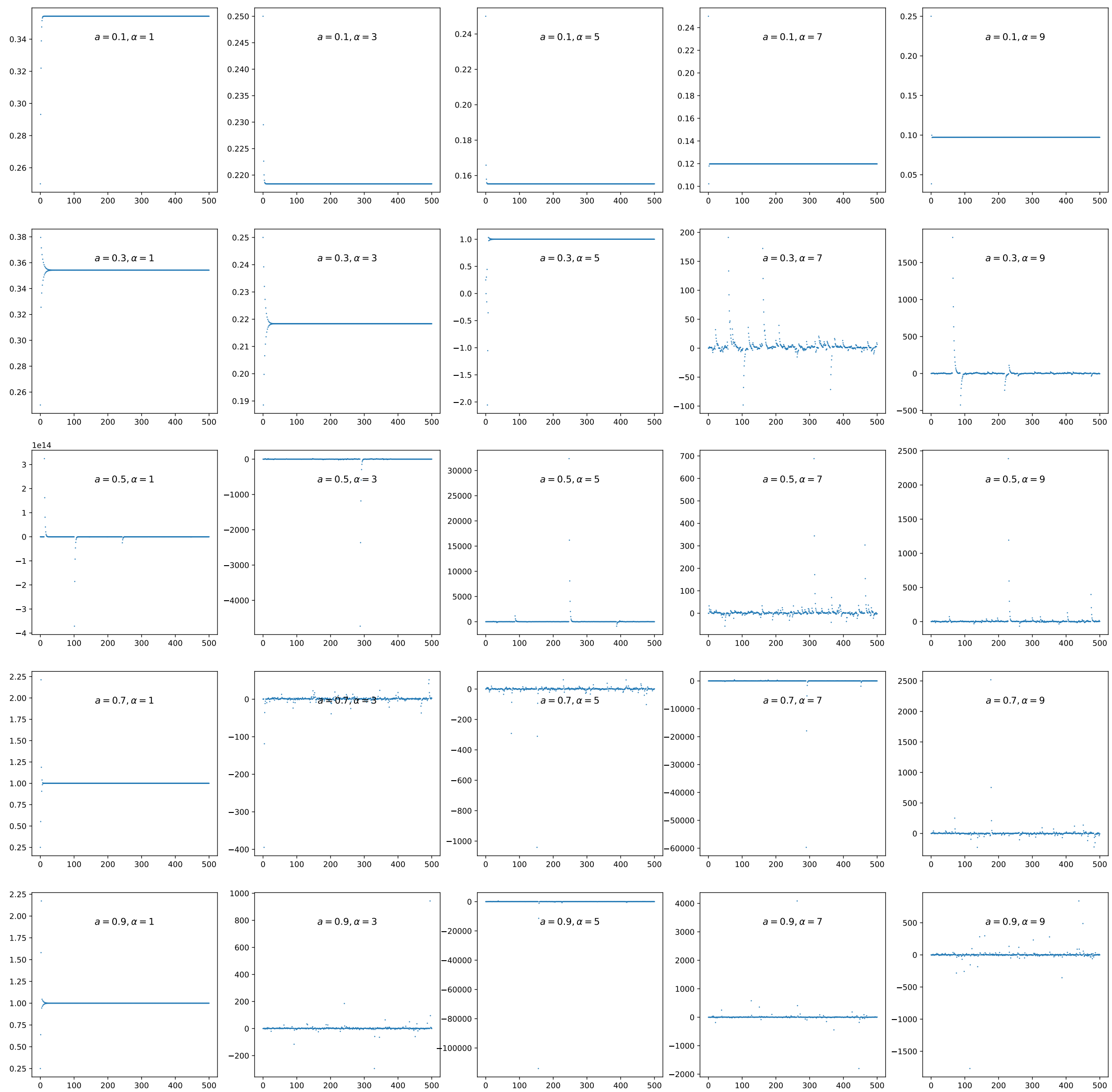
In our case, if recursive functions are such that $\epsilon = G(\epsilon)$, the G functions are, respectively

$$G_1(\epsilon) = (1-a)\epsilon + a(1 - \frac{\alpha(\tan(\pi\epsilon))}{\pi})$$
$$G_2(\epsilon) = (1-a)\epsilon + a(\frac{1}{\pi}\arctan(\frac{\pi(1-\epsilon)}{\alpha}))$$

We can check analytically by computing the derivative in the fixed point, or graphically see which of the choices of a yield to solutions that respect the condition in the neighbourhood of the fixed point.

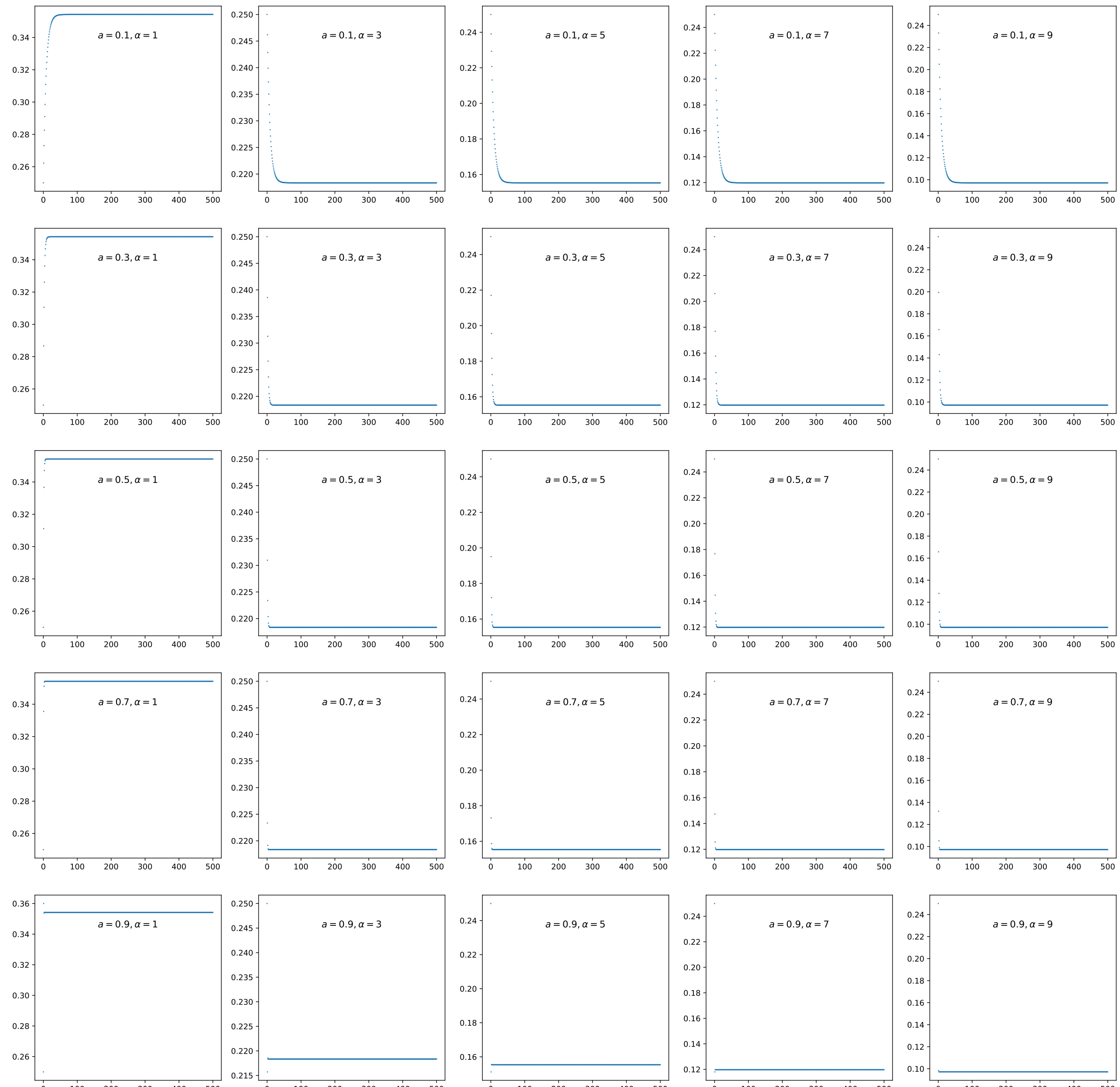


We can perform a grid search over the parameter space of the f_1 convergence. The grid represents convergence plots, going from the left to the right with α values 1, 3, 5, 7, 9 and going from the top to the bottom with a values 0.1, 0.3, 0.5, 0.7, 0.9.

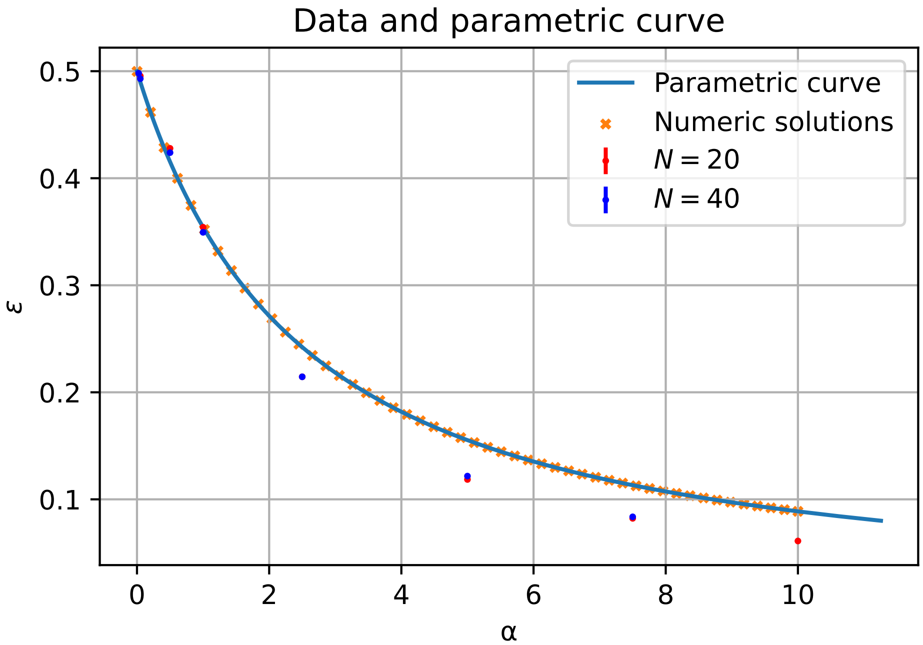


As a grows, we can see the onset of non-convergent behaviour on progressively lower values of α.

We can check by doing a grid search over the parameter space that f_2 works much better. See the following grid: we have convergence for all the α's, for all settings of a. The grid represents convergence plots, going from the left to the right with α values 1, 3, 5, 7, 9 and going from the top to the bottom with a values 0.1, 0.3, 0.5, 0.7, 0.9.


$a = 0.1, \alpha = 1$
$a = 0.1, \alpha = 3$
$a = 0.1, \alpha = 5$
$a = 0.1, \alpha = 7$
$a = 0.1, \alpha = 9$
$a = 0.3, \alpha = 1$
$a = 0.3, \alpha = 3$
$a = 0.3, \alpha = 5$
$a = 0.3, \alpha = 7$
$a = 0.3, \alpha = 9$
$a = 0.5, \alpha = 1$
$a = 0.5, \alpha = 3$
$a = 0.5, \alpha = 5$
$a = 0.5, \alpha = 7$
$a = 0.5, \alpha = 9$
$a = 0.7, \alpha = 1$
$a = 0.7, \alpha = 3$
$a = 0.7, \alpha = 5$
$a = 0.7, \alpha = 7$
$a = 0.7, \alpha = 9$
$a = 0.9, \alpha = 1$
$a = 0.9, \alpha = 3$
$a = 0.9, \alpha = 5$
$a = 0.9, \alpha = 7$
$a = 0.9, \alpha = 9$

Step 7 and 8: Use the chosen function to find the value of the function for many ϵ's and plot the final curves



The parametric curve and the numeric solutions agree perfectly, while the numeric solutions lie well below the formers.