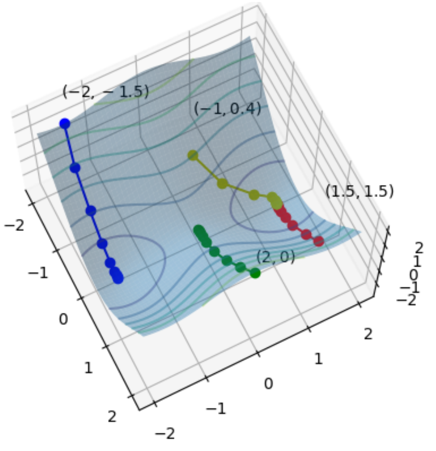
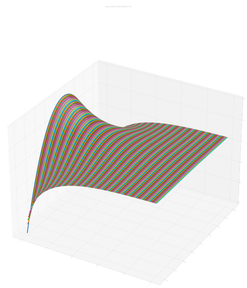


(a) SDM applied to $f(x, y) = -3\cos x + \cos^2 y$ at different initial points.



(b) Solution to L^2 gradient flow equation with 1D energy functional $\mathcal{E}(f) := \int_{-1}^{1} |\nabla f(x)|\, \mathrm{d}x$, which pe- nealises variation in function. Note that the solution converges to a function with no variation.

0.1 Steepest Descent to Gradient Flow Equation

For minimising a differentiable function $f : E \subset \mathbb{R}^n \to \mathbb{R}$, there is a well-known method known as **steepest descent method** (SDM)[**doi:10.1137/1.9781611974997.ch8**].

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha_k \nabla f\left(\mathbf{x}^k\right) \tag{1}$$

Starting from the initial input point $\mathbf{x}^0$, at each iteration, input points $\left\{\mathbf{x}^k\right\}$ move in the direction of "steepest" decrease, with specified step size $\alpha_k > 0$, reducing the value at evaluation of f. Note that in general, this method is not guaranteed to find the minimiser. On the other hand, convergence is guaranteed under certain assumptions, for example, convexity and L-smoothness with a certain choice of step size α_k.

Analogously, differential equation known describing the reduction process of a functional $F : \mathcal{F} \subset \mathcal{X} \to \mathbb{R}$ (where $\mathcal{X}$ is a function space) can be motivated. Starting from (1), replacing $\mathbf{x}^k$ by f_k and $\nabla f\left(\mathbf{x}^k\right)$ by $\operatorname{grad}_{\mathcal{X}} F\left(f_k\right)$

$$f_{k+1} = f_k - \alpha_k \operatorname{grad}_{\mathcal{X}} F\left(f_k\right) \tag{2}$$

Now think of f_k as "snapshots" at certain time $t = t_k$. Without loss of generality, let $\alpha_k \equiv 1$.[1] Dividing (2) by time step $\Delta t := t_{k+1} - t_k$, and taking the limit as $\Delta t \to 0$, we acquire the **gradient flow equation**[**YSC2021**].

$$\frac{\partial f}{\partial t} = -\operatorname{grad}_{\mathcal{X}} F(f) \tag{3}$$

where index k transforms to "time" variable t.

Note that grad of a functional is not defined yet. This depends on the function space $\mathcal{X}$ (eg. L^2, H^1, …) of interest.

[1]This is justified by taking a differrent time scale; essentially nondimensionalisation.