# Optimization with Surrogate Neural Network Functions

December 19, 2024

## 1 Introduction

The purpose of this code is to optimize a function defined by a surrogate neural network using CasADi's optimization tools. This document explains the mathematical formulation behind the code.

## 2 Mathematical Formulation

### 2.1 Objective Function

The optimization problem aims to minimize the entropy of a system of weights associated with the surrogate neural network. The entropy function is defined as:

$$H(\lambda) = -\sum_{i=1}^{n} \left(\lambda_i \log_2(\lambda_i + \epsilon) + (1 - \lambda_i) \log_2(1 - \lambda_i + \epsilon)\right), \tag{1}$$

where:

- $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ is the probability distribution.
- $\epsilon$ is a small constant added to avoid $\log(0)$.

This is implemented using the following steps:

1. Define $\lambda_{\text{min}}$ and $1 - \lambda_{\text{min}}$ with $\epsilon$ added.
2. Apply a smoothing function using `logsumexp` to stabilize computations.
3. Compute the entropy as a sum of the terms.

### 2.2 Surrogate Neural Network

The surrogate neural network is a function approximator trained to predict responses over the input space $(x, y)$. It is represented as:

$$z_k(x, y) = \text{surrogate\_func}(x, y, \text{parameters}). \tag{2}$$

The function surrogate_func models the system behavior and provides predictions for the optimization process.

## 3 Visualization of Results

### 3.1 Surrogate Neural Network Output

The surrogate neural network output corresponds to an array of the same dimension as the parameter array (representing the positions of target objects). The output is visualized to analyze the predictive behavior of the network.

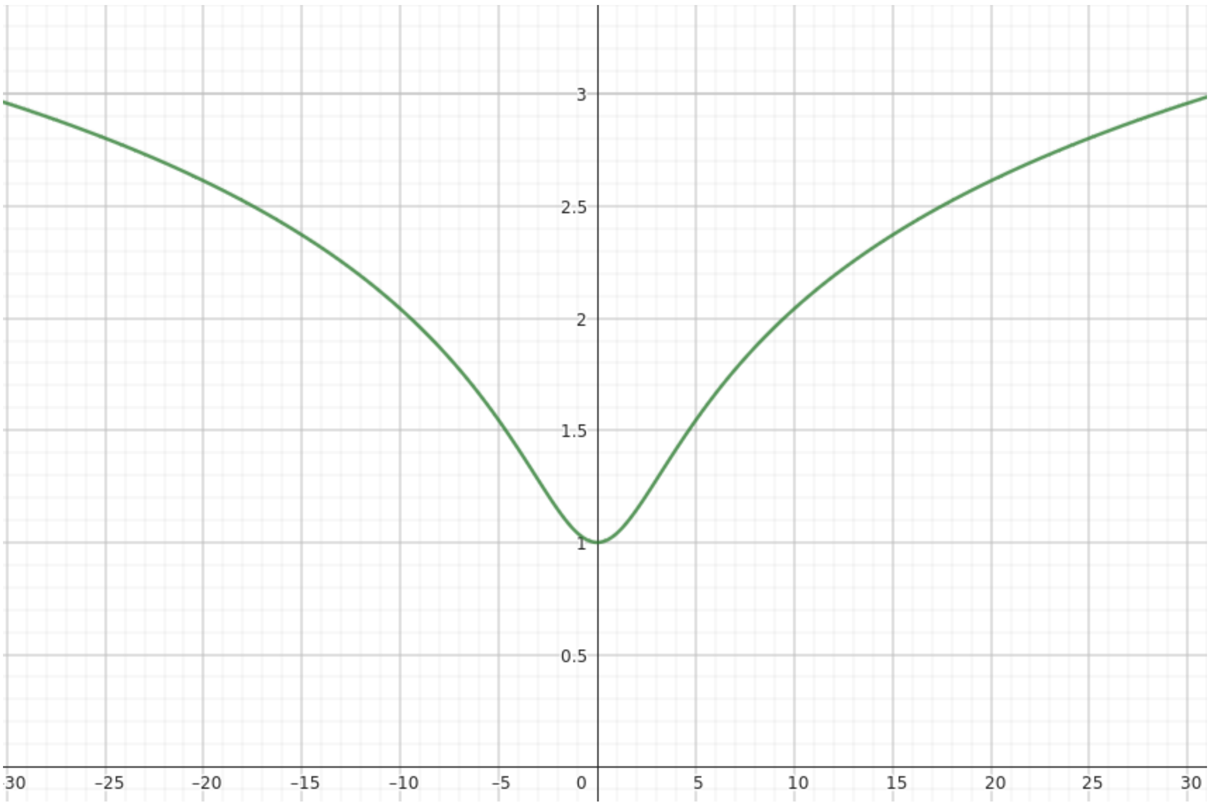
### 3.2 Weight Function for Avoiding Zero Gradient Areas

To avoid zero-gradient areas between objects of interest in the entropy function $H$, a weight function is applied. This weight function ensures that the optimization process avoids regions with no significant gradient, improving convergence.

### 3.3 Combined Function: Attractive and Bayesian Terms

The combined function adds the output of the attractive function and the Bayesian terms. The attractive function is defined as:

$$\text{attractive\_function}(x, y) = \max\left(-\left(1 + \log_{10}(0.001 \cdot f_Z(x, y) + 1)\right) \cdot \left(1 - 2(b_0 - 0.5)\right)^{-2}\right), \quad (3)$$



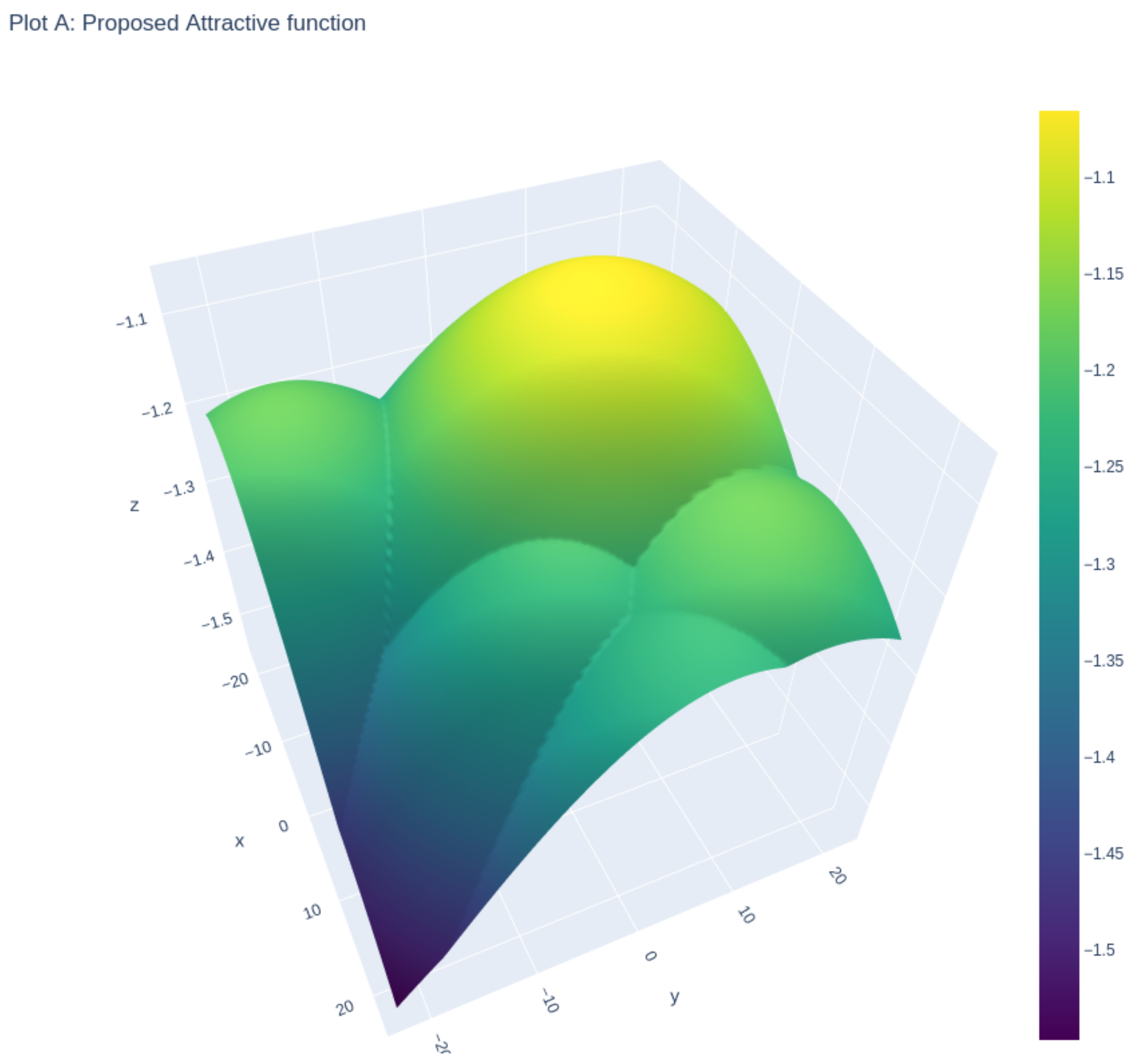

where $f_Z(x, y)$ is the distance-based function described symbolically and computed for target positions.

### 3.4 Entropy Output

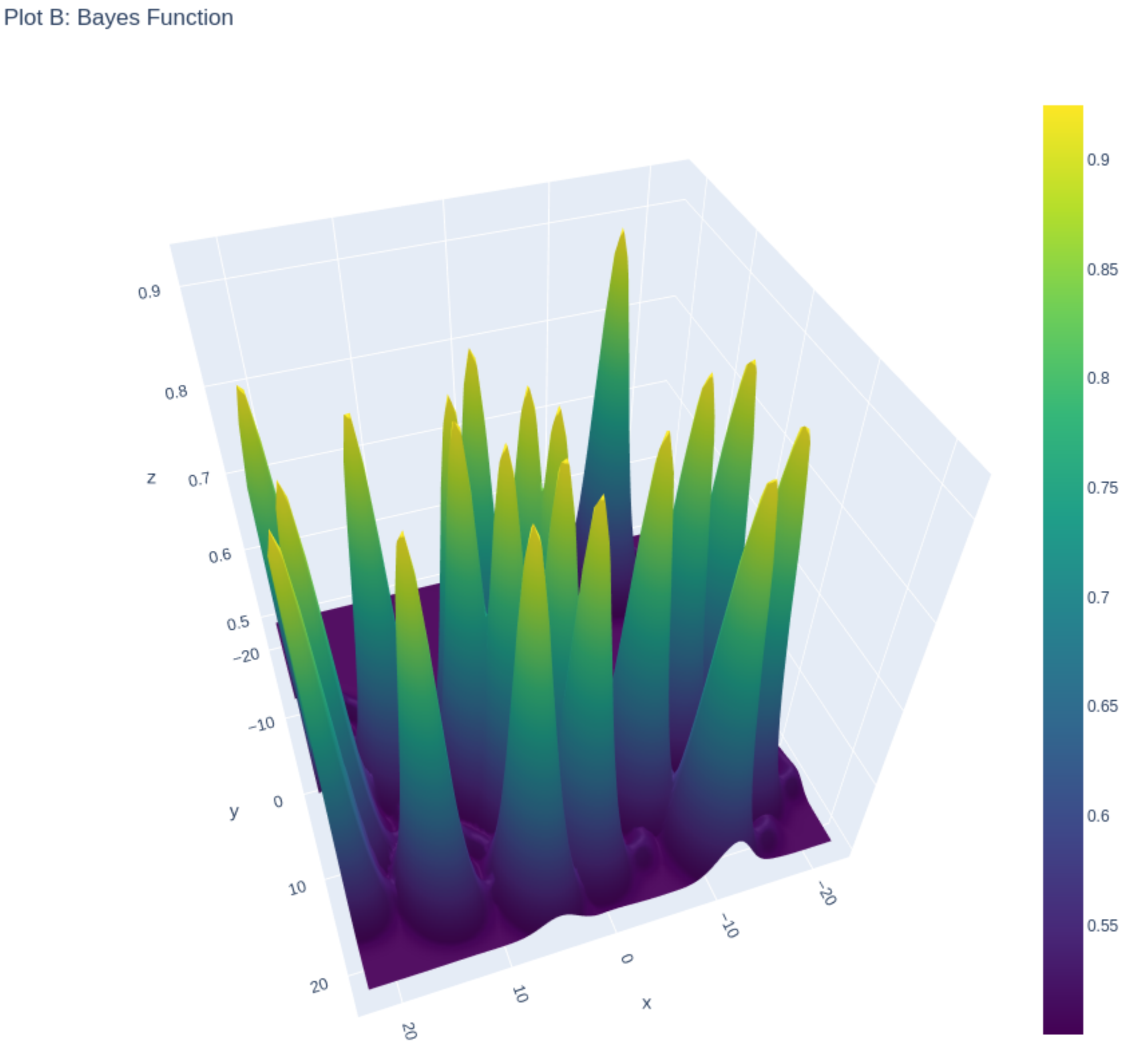
The entropy output is computed based on the updated weight function and is visualized to provide insights into the optimization process. This visualization highlights the regions of high entropy and their reduction over iterations.

# 4 Plots

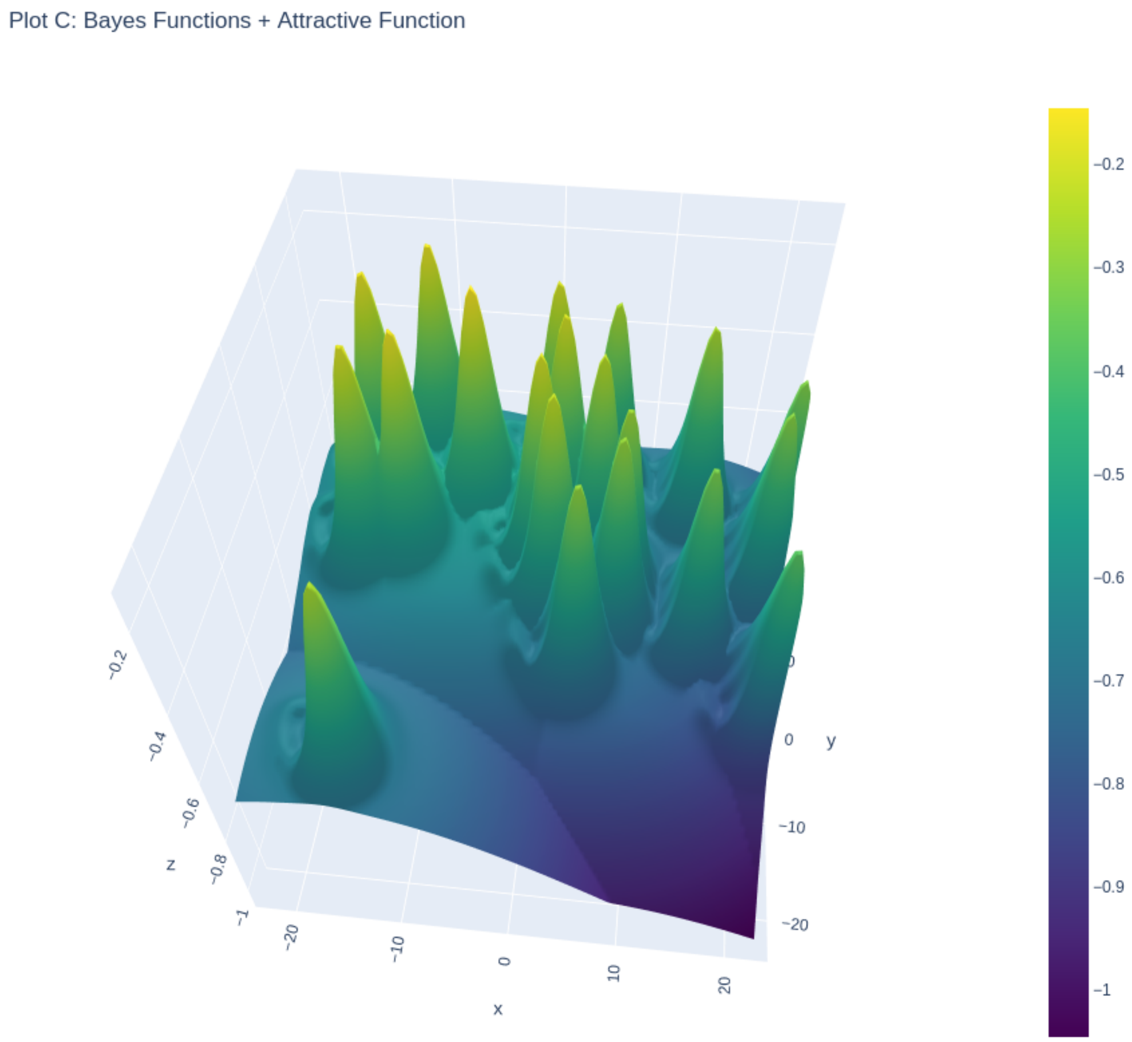
## 4.1 Weight Function



## 4.2 Surrogate Neural Network Output



### 4.3 Combined Attractive and Bayesian Function



### 4.4 Entropy Output