SURROGATE-ASSISTED MULTI-OBJECTIVE BAYESIAN OPTIMIZATION OF BIO-PRINTING PROCESS PARAMETERS

By

ADITYA RANE

Bachelor of Science in Mechanical Engineering
Mumbai University
Mumbai, Maharashtra
2018

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
MASTER OF SCIENCE
DECEMBER, 2023

SURROGATE-ASSISTED MULTI-OBJECTIVE BAYESIAN OPTIMIZATION OF BIO-PRINTING PROCESS PARAMETERS

Thesis Approved:

Dr. Akash Deep

Thesis Advisor

Dr. Srikanthan Ramesh

Dr. Chenang Liu

ACKNOWLEDGMENTS

Embarking on this academic journey and completing this thesis has been transformative. I owe the successful completion of this thesis to the unwavering support and guidance of key individuals who played pivotal roles in shaping my academic growth.

I would like to express my deepest gratitude to my advisor, Dr. Akash Deep, for the constant support, invaluable guidance, and continuous encouragement throughout the research process. His expertise and dedication have been valuable in shaping this thesis.

I am also indebted to the members of my thesis committee: Dr. Srikanthan Ramesh, for sharing his technical expertise in bioprinting, and Dr. Chenang Liu, for his insightful feedback throughout my research. Their knowledge added immense value to the development of this work. A special thanks to my labmate, Stephanie Hart, for her collaborative spirit, shared insights on CFD simulations, and her friendly attitude that made the research journey enjoyable.

My heartfelt appreciation goes to my family - Mr. Satish Rane, Mrs. Sarita Rane and Mr. Nandkumar Rane for their support, understanding, and encouragement. Their love and belief in me have been the pillars of strength that sustained me through this academic endeavor. I am thankful to all my friends who have made this academic endeavor not only intellectually fulfilling but also joyfully memorable.

Finally, I want to express gratitude to the teaching and non-teaching staff of the Department of Industrial Engineering and Management who may not be directly named here but have nonetheless contributed to my intellectual growth. Thank you all for being part of this academic journey.

Acknowledgments reflect the views of the author and are not endorsed by committee members or Oklahoma State University.

Name: ADITYA RANE

Date of Degree: DECEMBER, 2023

Title of Study: SURROGATE-ASSISTED MULTI-OBJECTIVE BAYESIAN OPTIMIZATION OF BIO-PRINTING PROCESS PARAMETERS

Major Field: INDUSTRIAL ENGINEERING AND MANAGEMENT

Abstract: 3D bioprinting is an additive manufacturing method that allows bio-mimetic fabrication of living cells through layer deposition of bio-ink. Tissue constructs resulting from such additive manufacturing process have the potential to emulate functional characteristics of native human tissue microenvironment. This emerging technology has various biological research applications mainly in regenerative medicine, in-vitro disease modelling and drug screening. Due to this fact, it could be possible to transplant damaged human organs and study drug responses to diseases. Extrusion-based bio-printing (EBB) is widely used for the process of biofabrication, provides a practical solution for the fabrication of tissue constructs. Its principle of operation is based on exerting pressure on bioink through the orifice of nozzle to form filaments or droplets. As a consequence of this operation, cells are damaged either due to mechanical forces or shear stress. Furthermore, the interplay of bioink composition with nozzle design parameters has a detrimental effect on living cells. Unlike non-biological printing, 3D bioprinting has additional complexities of cell viability and post-printing functionality.

In bioprinting, design optimization of printer parameters, to minimize the exerted shear stress on wall of nozzle and maximizing the survival as well as functional characteristic of cells is crucial. An experimental design approach is time consuming and cost intensive. This thesis introduces data-driven Gaussian Process Regression Multi-Objective Bayesian Optimization (GPR-MOBO) framework. The approach optimizes five design parameters including nozzle and bio-ink rheology: outlet diameter (d), maximum and minimum viscosity (η), power law (n) and flow consistency index (K). Initial experiments are conducted with two-level factorial design by varying ranges of design inputs. Using CFD simulations of 18G and 25G nozzle in Ansys FLUENT three fluid shear responses are extracted at the X-Y-Z location of the internal fluid shear profile. The proposed GPR-MOBO framework combines the principle of black-box modeling (a function without closed-form relationship) using the Gaussian process and integrates sampling and optimization using Bayesian optimization. On account of integration, the experimental approach becomes sequentially adaptive and determines the Pareto optimal sampling point in 21 iterations following preliminary 24 runs. The proposed framework implements adaptive sampling method for creating an iterative framework and optimally selecting process inputs to minimize shear stress and maximize cell functionality.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Chapter I, outlines the complexity of the bioprinting process and the significance of the machine learning model to address such problems. Section 1.1 focuses on complexities encountered in extrusion-based bioprinting. Section 1.2 delineates the importance of optimization within this context and expounds on the GPR-MOBO framework as an active learning strategy. In Section 1.3, our discourse centers on the precise research objectives encompassed within this thesis.

1.1 Complications in tissue engineering and extrusion-based bioprinting

Over the last decade, the global demand for organ transplantation has exponentially increased due to successful post-transplant outcomes [61]. However, this rapid surge in demand exacerbated the critical issue of organ shortages. Similarly, there is an advancement in generating disease models in mice by genetic modification nevertheless, this model fails to imitate human conditional response to drug screening [29]. This has become evident through notable research in the field of regenerative medicine and in-vitro disease modeling with the aid of tissue engineering [61]. An amalgamation of biology and material science, this domain emerges as a promising solution to reconstruct synthetic tissue and organs with the potential of creating intricate tissue structures with remarkable precision and functionality [26].

Bioprinting, an additive manufacturing approach, is commonly used in tissue engineering. This can be implemented with three distinct mechanisms droplet-based (DB), laser-based, and extrusion-based bio-printing (EBB) [16]. Among all, EBB is widely used due to its ability to fabricate with bio-inks having high polymer concentration and [58]. Unfortunately,

the major challenge in EBB is cell viability because a higher percentage of cells are subjected to mortality due to cell stresses induced during bio-ink extrusion through the nozzle needle and printing pressure [16][64]. As cell viability determines the magnitude of the cells that remain functional post-printing process, a critical factor contributing to cell functionality is wall shear stress in the nozzle which remains a challenge in extrusion bioprinting [34][38]. As shown in Figure 1, describe the desirable characteristics of non-Newtonian shear thinning bio-ink. Figure 2 depicts the underlying relationship between shear stress and print parameters which generally results in myriad trade-offs between printing parameters and bio-ink rheology. Some of them include but are not limited to pressure, outlet diameter, velocity, viscosity, non-Newtonian power law, and flow consistency index. Considering the printing

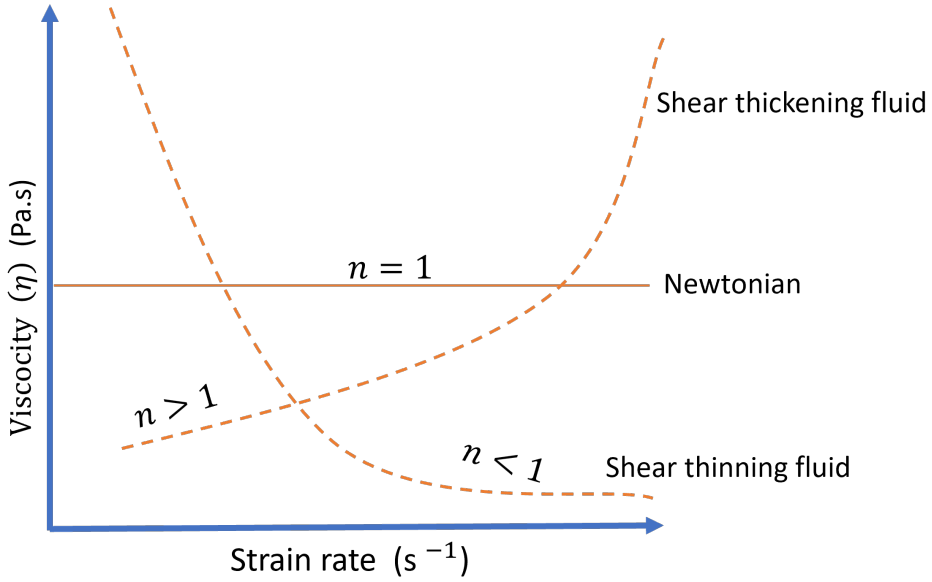

Figure 1: Characteristics of bio-ink rheology

parameters perspective significant contributing factors to cell damage are inlet pressure and gauge diameter of the nozzle, increasing the pressure and reducing the nozzle diameter will increase the shear stress. Similarly, a higher value of inlet velocity can reduce cell viability. Shear stress increases at higher pressure, velocity, and smaller outlet diameter. On the other hand, considering the bio-ink rheology, high viscosity will engender severe cell damage. Furthermore, as shown in figure 1, the power law index shows (n) the fluid behaviour, a value

$n < 1$ results in a greater degree of shear thinning whereas $n > 1$ results in shear thickening. Ideally, the desirable property of bio-ink should possess shear thinning behavior i.e. shear rate increases with a decrease in viscosity that will prevent nozzle blockage and reduce shear stress. Finally, the preliminary viscosity of the bio-ink is related to the flow consistency index which is related to the extrudability of the bio-ink i.e. the power required to force fluid through nozzle orifice, lower value indicates higher extrudability. Achieving the balance between printing and bio-ink rheology is crucial to reduce the shear stress and increase the number of living cells post-printing.

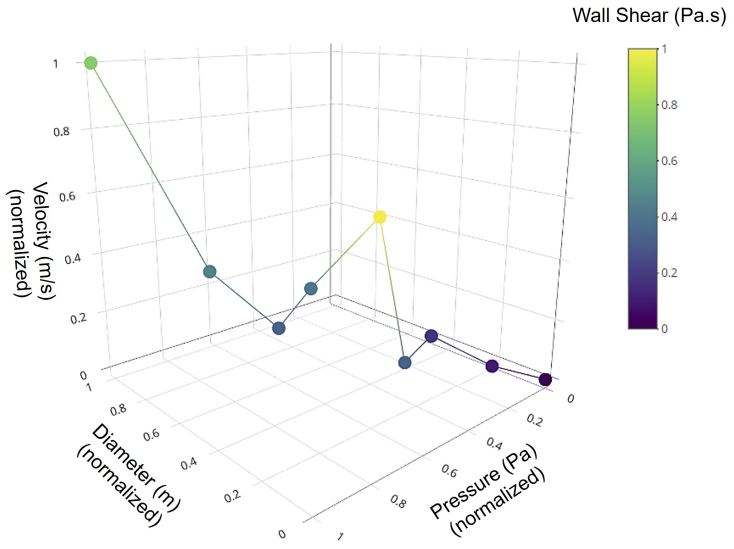

Figure 2: Printer parameter influence on cell viability based on the data source of nozzle design [33]

1.2 Optimization strategies for bio-printing

Printer parameter optimization and evaluation of the aforementioned critical issue are traditionally expensive and time-consuming due to trial-and-error approaches. However, the advent of data-driven methodologies can tackle this issue with minimum observations of experimental data that can help to select the process parameters and consider the complex interplay of several factors affecting the tissue construct properties [30]. Machine learning models have gained wide popularity in the context of additive manufacturing, specifically

3D printing due to the potential of optimizing the printing parameters. Nonetheless, limited applications are found in 3D bioprinting where optimizing the bio-ink rheology and printer parameters is significant [6]. As Figure 3, depicts the complexities involved in modeling the parameters without a closed-form relationship to fabricate a 3D cellular matrix, tuning such parameters could be a time-consuming or rather costly undertaking. Negligence of such factors will lead to a clogged nozzle, compromised mechanical strength, or sedimentation of non-functional cells. Bio-ink is composed of living cellular entities and therefore achieving the precise and effective deposition of living cells is critical.

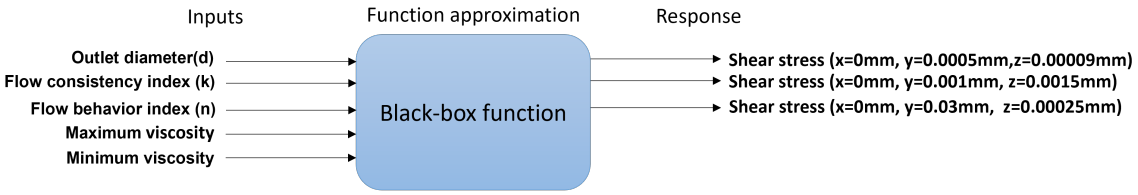

Figure 3: Process parameters modeling

Researchers have explored an alternative approach to address the challenges in 3D bioprinting optimization such as the classical design of experiments (DOE), Latin Hypercube sampling, min-max, and space-filling design [21]. Despite these efforts, the main impediments of these methods are sampling patterns pre-established prior to conducting the experiments. Furthermore, it fails to adapt to newly emerging features that arise during the experiment. In contrast to classical experimental design, the adaptive sampling method integrates the Gaussian process and acts as a non-parametric surrogate model. It establishes the prior distribution over the process parameters and fluid shear responses and selects the function that best fits the given set of data points. This circumvents the complexity of modeling function that exhibits non-linear behaviour [59]. Bayesian optimization (BO) serves as a sample efficient method for the function that exists without closed-form relationships – known as black-box functions. Mean and variance information from the GP is helpful for BO, to establish a balance between two strategies of exploration (unexplored design space) and exploitation (regions with uncertainty) in order to determine the next promising process

parameters. This framework is iterated in a loop until the stopping criteria are reached i.e. either budget assigned to experiments or variance becomes negligibly small in magnitude.

1.3 Research Objectives

As the domain of bio-printing reveals the notable absence of sequential experimental design our research objectives of the proposed framework are as follows:

1. Implement a data-driven optimization approach that utilizes the GPR-MOBO framework to systematically improve the bio-printing process for tissue engineering.

2. Investigate the combined interplay between bio-ink rheology and print parameters on cell functionality during the 3D bio-printing process.

3. The fundamental objective is to diminish internal fluid shear stress, consequently leading to a reduction in wall shear stress.

4. Replacing the experimental design approach with adaptive sampling that will sequentially optimize the process parameters ultimately improving the efficiency and effectiveness in the field of bio-printing.

CHAPTER II

LITERATURE REVIEW

In engineering research problems, computer simulation systems such as Computational Fluid Dynamics (CFD) and Finite Element Analysis (FEA) are extensively used to replicate the response of the physical system however results from such systems are deterministic without stochastic effects. Moreover, classical experimental methods are non-adaptive as they follow fixed sampling patterns [44][54]. In the following literature review, we delve into several classical experimental designs in section 2.1, computer simulation models in the design of experiments are outlined in section 2.2. Section 2.3 emphasizes the application in the field of material design and additive manufacturing.

2.1 Classical methods in design of experiments (DOE)

Experimental design plays a crucial role in scientific research. Conventional design methodologies are as follows. In the random design of the experiment, as the name suggests experimental units are assigned to different groups randomly to eliminate systematic biases during allocation [23]. Randomized designs are generally suitable for larger data sets however, their efficacy is limited for smaller data sets [8]. Additionally, the effectiveness of the experimental results depends on the number of replications. Higher replication reduces the variability in results and it's the only way by which researchers can conclude with fidelity [7]. Another approach to DOE is factorial experimental design has discrete possible levels with two or more factors, and units of experiments take all combinations across each factor [14]. Major impediments are at least two or more factors required and complexity increases with several factors. Moreover, sampling is done at orthogonal corners of the design space ignoring the

non-linearity of the model. In response to these challenges, alternative approaches have been developed such as Latin hypercube sampling (LHS) exhibits random probabilistic dispersion and max-min design is deterministic and allows dispersion in terms of relative distance. Both are of space-filling design and provide uniform coverage over the entire design space. Unlike random sampling, new points are generated independently without considering previously generated sample points [50]. Orthogonal arrays proposed by Genichi Taguchi are widely recognized for improving the quality of the process by considering the interplay of process parameters and their effect on the mean and variance of performance characteristics [57]. Instead of testing all combinations in the case of factorial design. Taguchi methods test pairs of experiments to reduce experimentation time. Nonetheless, the efficiency of this method is limited because all the combinations are not tested, and the interaction between parameters is ignored. Collectively, classical methods have pre-established sampling patterns, and they are non-adaptive which may not be well-suited for dynamically evolving features during the experiment. Furthermore, they are restricted in determining the next sampling location that converges the experiment to global optima therefore they are less cost-effective.

2.2 Response surface methodology (RSM)

Response surface methodology is a statistical tool that investigates the relationship between multiple independent variables and one or two dependent variables. Depending on the approximation function RSM are categorized into the first-order model, second-order model, and mixture model. Box and Wilson (1951) developed this approximation method for process optimization in the chemical industry highlighting its significance in experimental design. They asserted that the methodology is an approximation and lacks precision [4][36]. Nevertheless, this approach is not without its limitations. Raissi et al. (2009) conducted a comprehensive study using RSM in a multi-objective optimization design problem consolidating multiple responses as a single composite solution [47]. This resulted in a compromise solution neglecting the trade-off between the responses in multiple objective problems. Fur-

thermore, Sinha et al. (2012) developed a predictive model for the optimization of the dye extraction process using RSM. Their model is subjected to higher order complexity resulting in a lack of fit that changes prediction results that is, prediction results are compromised when the number of input parameters are increased [41][51][52]. The RSM model lacks robustness for a higher number of observations because it is non-parametric, meaning it has a predetermined structure. This characteristic makes it unfavorable for large-scale optimization [24][28][40]. The potential of RSM is limited in adaptive experiments because it emphasizes local search over global search [3][27]. In this approach, the selection of the starting point is crucial for initiating the search. The procedure involves measuring the response to various experimental conditions along the path of the steepest ascent. Subsequently, the best solutions are iterated based on the condition with the highest measured response along the gradient path. Consequently, RSM is focused on finding the next best solution in proximity to the initial starting point without exploring the entire design space. Additionally, the reproducibility of results becomes questionable if there is variation in the selection of the starting point.

2.3 Computer simulation models in the design of experiments.

Chand et al. (2022) proposed a cell viability assay by investigating the process parameters of a bioprinter using computer simulation on shear wall stress exerted in the nozzle to optimize the process parameters [6]. Additionally, they examined the effects of printing speed on the thread profile in tandem with outlet velocity and pressure, and they highlighted the interplay of the nozzle geometry, nozzle diameter, inlet pressure, and bio-ink rheology on maximum wall shear stress and cell viability. The variables under observation were assumed to be independent and a second-order polynomial model was deployed to perform computer simulations. They asserted the usefulness of the computer simulation model for reducing the number of experimental iterations nonetheless, this model suffers from drastic changes in flow rate predicting negative values of shear stress.

Lemarié et al. (2021) demonstrated that cell viability reduces as the magnitude of shear wall stress increases by investigating the impact of the bio-printing process parameters on the percentage amount of living cells and the quality of cells associated with biological compositions [30]. The prediction model for the underlying cells is a partial least square (PLS) approach and principal component analysis is executed to identify impacting parameters and the metric to measure model accuracy is the coefficient of determination. The model correctly predicts the living cells percentage with modest prediction error, however, the predictive models exhibited limited accuracy for the quality of cells as indicated by the coefficient of determination values falling below 0.7.

The aforementioned findings provide valuable insights into computer simulation models that fail to provide accurate results for complex systems specifically when higher dimensions are involved, and evaluations are expensive and time-consuming.

2.4 Applications in the field of material design and additive manufacturing.

Liu et al. (2022), proposed a non-parametric framework for material design and process optimization in additive manufacturing with the fused filament fabrication (FFF) process [32]. The aim is to integrate statistics, machine learning (ML), and optimization techniques to model the composition of process-property relationships (process parameters and graphene composition with surface quality values) and optimize them with a minimum number of experiments. This manuscript introduces single-objective Bayesian optimization with multivariate Gaussian distribution as a probabilistic model along with expected improvement as an acquisition function. This utility function balances out exploration, where prediction uncertainty is high, and exploitation when Gaussian process regression (GPR) predicts high objective value to iteratively scrutinize the search domain for the best surface finishes within the fewer experiments to save time and cost for nanocomposite design.

Olofsson et al. (2021) study was motivated by selecting the design inputs for multi-objective space having conflicting objectives in tissue engineering and regenerative medicine where the

objective is to maximize the neo-tissue growth (expensive to evaluate black box function) and minimize the cost of neo-tissue production associated with bioreactors (analytically cheap to evaluate) [43]. This manuscript introduces the probabilistic Pareto frontier (an extension for Bayesian MOO) with EMmI (expected max-min improvement) as an acquisition function approach to find the solution for multi-objective optimization problems with conflicting objectives and demonstrates the convergence with a minimum number of function evaluations that ensures a measure of confidence (uncertainty) in the selection of optimal points in unexplored regions of variable space. This method is compared with other MOO methods; genetic algorithms and random search that don't provide confidence in selecting the unexplored optima and additionally require a maximum number of function evaluations. Ruberu et al. (2021), Bayesian optimization along with the Gaussian process as surrogate model algorithms was applied to attain the goal of optimal printing parameters with a minimal number of experiments [49]. The primary focus of this study was on the printability to achieve 3D constructs with accurate geometric reproduction. The research was conducted with GelMA and GelMA/HAMA bio-inks with varied concentrations and six different bioprinter parameters namely (bio-ink reservoir temperature, extrusion pressure, print speed, and the platform temperature), and response from the black box function was a scoring system with overall score "0" demonstrates optimal printing condition however, the scoring system is based on the visual assessment and comparing the print with a standard construct which may result into human error. This could oversimplify the complexities of tissue construct quality and functionality which can be influenced by other factors.

Tian et al. (2021), study addresses the novel approach of EBB parameter optimization through machine learning models (ML), They investigate both regression and classification models to predict outcomes such as cell viability and filament diameter for composite bio-inks (gelatin and alginate) [59]. However, the robustness of the ML model is questionable when trained on datasets that include other assays targeting the same cellular aspects. The model was trained on data from published literature, and its accuracy appears to decline

as the size of the training dataset is augmented. This degradation in accuracy becomes evident through higher prediction percentage errors (MSE) when comparing experimental values with the complete datasets, undermining the goal of minimizing experimentation.

Solomou et al. (2018), proposed a closed-loop multi-objective material discovery based on Bayesian optimization experimental design (BOED) in realization with sequential executing optimal selection policy (OES) [54]. The suggested framework explores the design space to achieve the target properties that satisfy more than one objective. The Gaussian process regression model is used to surrogate the response of computational experiments. In the selection policy (OES) expected hypervolume improvement (EHVI) is used as an acquisition function for probing the material design space They compared the OES policy with pure random experiment selection (PRES) and pure experiment exploration selection (PREES), and the results suggest that desired properties of materials are achieved with a minimum number of experiments and preordained experimental budget with BOED in comparison with PRES and PEES where former is randomly dispersed in objective space and latter performs the selection in a specific region.

Park et al. (2018), developed CFD-MOBO the design of the chemical reactor to solve the CFD-associated design problems [45]. This framework was applied to minimize the power consumption and maximize the gas hold-up in the reactor tank along with six design variables. They suggest BO is suitable for the function that requires expensive black-box evaluations due to its ability to explore the search regions that are likely to be optimum. Results obtained from the Pareto front were compared with the reference design, it turns out that optimal design shows better results that achieve the required objective as compared with the reference design. Furthermore, the automation in the process of selection of valid inputs is obtained by forming a link between CFD and MOBO. Their study states the generalization of this approach to design problems that require a significant number of CFD runs.

CHAPTER III

METHODOLOGY

3.1 Black-box modeling using Gaussian process

A Gaussian process is a stochastic process such that a collection of a finite set of random variables has a multivariate normal distribution [20][48]. The linear combination of the random variables is also normally distributed where $\mathbf{x} \in \mathbb{R}^q$, represents the design vector from 'q' dimensional input space $\mathcal{X}$, $\mathbf{x} \in \mathcal{X}$. The specification of the Gaussian process entirely relies on the mean function $\mu(\mathbf{x})$ that determines the expected value of the distribution and covariance function $k(\mathbf{x}, \mathbf{x}')$ (also known as kernel function) describes how different random variables are correlated where $\mathbf{x}$ and $\mathbf{x}'$ are two distinct random variables [12].

$$f(\mathbf{x}) \sim \mathcal{GP}((\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \tag{3.1.1}$$

The mean and covariance function for $f(\mathbf{x})$ i.e., the value of the function at location $\mathbf{x}$ can be written as follow:

$$\mu(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})] \tag{3.1.2}$$

$$k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \mu(\mathbf{x})(f(\mathbf{x}') - \mu(\mathbf{x}')] \tag{3.1.3}$$

Gaussian distributions possess a valuable algebraic characteristic – they maintain their Gaussian nature when subjected to conditioning and marginalization [11][19]. This property entails that any distributions resulting from these operations will also retain the Gaussian form.

Marginalization and conditioning are the two special properties of the Gaussian distribution. Former extracts partial information from the probability distribution that consists of multiple variables therefore making it possible to focus on variables of interest. On the other hand, the latter entails the probability of one variable given another. Similarly, it preserves the Gaussian nature of the distribution by modification in Gaussian distribution [19].
Marginalization and conditioning over the Gaussian distribution can be given as a normal probability over the subspace of random variable vectors (see Appendix A.1). Conditioning is significant in widening the scope of Gaussian process modeling and Bayesian inference for making the decision on the basis of available information [19].

3.1.1 Kernel functions in Gaussian process

In the Gaussian process, $\mu = 0$ for the simplicity of conditioning this is also referred to as centering of mean often essential in normalization [12]. The underlying challenge lies in setting up a covariance function also known as (kernel function or covariance kernel) which specifies the nature of the distribution we want to predict along with its shape. The positive definite kernel function has two inputs $\mathbf{x}$ and $\mathbf{x}'$ which determine the degree of similarity as a scalar quantity. GPR models use a kernel function to define prior information about the correlation of function evaluated on each object.

$$\Sigma = \mathrm{Cov}(\mathbf{X}, \mathbf{X}') = k(\mathbf{x}, \mathbf{x}')$$

Different kinds of kernels used in the Gaussian process are mentioned below:

Square Exponential kernel

A square exponential kernel also known as a radial basis function (RBF) kernel is a stationary kernel i.e., the Euclidean distance between the two points $\mathbf{x}$ and $\mathbf{x}'$ and correlation only

depends on the location of the two points [12].

$$\sigma^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\ell^2}\right) \tag{3.1.4}$$

where ℓ^2 and σ^2 are hyperparameters that regulate the smoothness and scale of the kernel respectively.

Periodic kernel

Similar to SE, the periodic kernel is also a stationary kernel that depends upon the relative displacement of two input points. Specifically, useful when modeling functions with periodic behavior where parameter p determines and regulates the distance between each repetition [12].

$$\sigma^2 \exp\left(-\frac{2}{\ell^2}\sin^2\left(\frac{\pi\|\mathbf{x}-\mathbf{x}'\|^2}{p}\right)\right) \tag{3.1.5}$$

Linear kernel

Linear kernel is the non-stationary kernel that generates different predictions if data is transversed parameter c determines the point of intersection for all the functions.

$$\sigma^2(\mathbf{x}-\mathbf{c})(\mathbf{x}'-\mathbf{c}) \tag{3.1.6}$$

Matern kernel

In Gaussian process modeling, the matern kernel provides finer control over the different degrees of smoothness [20].

$$k_\nu(r) = \frac{\Gamma(\nu)}{2^{\nu-1}\Gamma(1-\nu)}\left(r\sqrt{\frac{2\nu}{\ell}}\right)^{\nu} K_\nu\left(r\sqrt{\frac{2\nu}{\ell}}\right) \tag{3.1.7}$$

In the above equation, k_ν is the bessel function, $\Gamma(\nu)$ is the gamma function, r is the Euclidean distance between two points and ν is the shape parameter that controls the degree

of smoothness which takes positive real value such that smaller value result in smoother function.

3.1.2 Gaussian Process Regression

Supervised machine learning models like GPR also known as kriging are derived from the probability distribution obtained from the Gaussian process, which plays a crucial role in establishing the kriging equations [62]. Kriging equations are essential because they provide insights into spatial correlation and uncertainty within the data [48]. The ability to perform estimation for specific locations by calculating the weighted average of known values at surrounding locations which makes GPR, is well-suited for both classification and regression tasks.

Now, the Gaussian process regression model generates the prediction by integrating the prior information (kernel) while offering means to quantify the uncertainty associated with predictions [62]. This is elaborated with single-output, section 3.1.2, and multi-output 3.1.3 Gaussian regression model [20][65].

3.1.3 Single-output multivariate Gaussian Process Regression (SO-GPR)

The multivariate single output Gaussian regression function is given as:

$$p(f|\mathbf{X}) \sim \mathcal{N}(f|\mu, K) \tag{3.1.8}$$

In GPR, we work with a set of inputs represented as $\mathbf{X} = [\mathbf{x_1}, \cdots, \mathbf{x_n}]$, each having corresponding function values $f = [f(\mathbf{x_1}), \cdots, f(\mathbf{x_n})]$ and $\mathcal{D}_n = [\mathbf{X}_n, Y_n]$ are associated training sets. Moreover, we introduce a mean function $\mu = [m(\mathbf{x_1}), \cdots, m(\mathbf{x_n})]$ and define a kernel function $\mathbf{K}_{i,j} = k(\mathbf{x_i}, \mathbf{x_j})$. $m(\cdot)$ represents the mean function, and $k(\cdot, \cdot)$ is a positive definite kernel function [48][62].

When we have no prior observations, the mean function is set to a $\mu(\mathbf{X}) = 0$, especially when working with normalized data that has a zero mean [62]. The smoothness of the

function is determined by $\mathbf{K}$ [48]. In addition to this, if the kernel considers two data points, $\mathbf{x_i}$ and $\mathbf{x_j}$, to be similar, it implies that the outputs of the function at these points, $f(\mathbf{x_i})$ and $f(\mathbf{x_j})$, are expected to be similar as well [21][48][62]. Now, let say we have to make predictions $f(\mathbf{X}_*)$ at new point $\mathbf{X}_*$ then joint probability distribution of f and f_* for noise free case is represented as:

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu(\mathbf{X}) \\ \mu(\mathbf{X}_*) \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{bmatrix} \right) \tag{3.1.9}$$

The posterior prediction of the mean and covariance function can be obtained by taking the conditional distribution of f_* can be written in terms of f as mentioned in equation (3.1.10). Equation (3.1.11) and (3.1.12) represent predictive information from Gaussian distribution in closed-form representation. Where $\mathbf{K}$ is the correlation matrix, $\mathbf{K}_*$, and $\mathbf{K}_{**}$ are covariance and variance matrix respectively.

$$p(f_*|f) \sim \mathcal{N}(\mu^*, \Sigma^*) \tag{3.1.10}$$

$$\mu^* = \mu(\mathbf{X}_*) + {\mathbf{K}_*}^T \cdot \mathbf{K}^{-1}(Y - \mu(\mathbf{X})) \tag{3.1.11}$$

$$\Sigma^* = \mathbf{K}_{**} - \mathbf{K}_*^T \cdot \mathbf{K}^{-1} \cdot \mathbf{K}_* \tag{3.1.12}$$

3.1.4 Multi-output multivariate Gaussian Process Regression (MO-GPR)

Multi-output Gaussian regression has become prevalent in various engineering disciplines where the essence lies in the prediction across all the outputs [5][31]. Such a type of regression model is also called a multi-variate kriging which focuses on multi-objective learning by leveraging the fundamental presumption that there exists some form of correlation among the outputs [5][60].

Similary to SO-GPR, the output of the MO-GPR is assumed to follow the Gaussian process with modification in the multi-output covariance matrix ($\mathbf{K}$) [54][60]. Multi-output model-

ing for 'n' queried points and 'M' inputs, the kriging equations can be obtained as follows.

The input matrix $\mathbf{X}$ consisting of 'M' input variable for 'n' queried points and $\mathbf{Y}$ represent the output values for the given input with 'N' number of objectives. The training dataset for the input-output pairs is provided as $\mathcal{D}_n = \{\mathbf{X}_n, \mathbf{Y}_n\}$. In matrix $\mathbf{X}$ each row corresponds to input vector $\mathbf{x}^i$. Similarly, each column of matrix $\mathbf{Y}$ corresponds to output vector $\mathbf{y}_i$. Note, the i^{th} superscript and subscript denotes the row and column vector respectively. Now, $Y_{ij} = f_j(\mathbf{x}^i)$ provides the value of j^{th} objective function depending on input $\mathbf{x}^i$.

Furthermore the constant mean function can be given as $h(\mathbf{a}; \boldsymbol{\theta})$ and $\mathbf{K}(\mathbf{A}, \mathbf{A}^{'}; \boldsymbol{\theta})$ is the covariance function for the combination of arbitrary matrices $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots \mathbf{a}_l \end{bmatrix}$, $\mathbf{A}^{'} = \begin{bmatrix} \mathbf{a}_1^{'} & \mathbf{a}_2^{'} & \cdots \mathbf{a}_p^{'} \end{bmatrix}$ where l and p represent row and column respectively. Furthermore, hyperparameter $\boldsymbol{\theta} = \begin{bmatrix} \alpha_1 & \alpha_2 & h(\cdot) \end{bmatrix}^T$ for square exponential kernel function $k(\mathbf{a}, \mathbf{a}^{'}, \theta)$ where α_1, α_2 and $h(\cdot)$ are scale, length, and mean function respectively.

$$\mathbf{X} = \begin{bmatrix} X^{11} & X^{12} & \cdots & X^{1M} \\ X^{21} & X^{22} & \cdots & X^{2M} \\ \vdots & \vdots & \ddots & \vdots \\ X^{n1} & X^{n2} & \cdots & X^{nM} \end{bmatrix}, \mathbf{Y} = \begin{bmatrix} Y^{11} & Y^{12} & \cdots & Y^{1N} \\ Y^{21} & Y^{22} & \cdots & Y^{2N} \\ \vdots & \vdots & \ddots & \vdots \\ Y^{n1} & Y^{n2} & \cdots & Y^{nN} \end{bmatrix}, \mathbf{K} = \begin{bmatrix} K^{11} & K^{12} & \cdots & K^{1p} \\ K^{21} & K^{22} & \cdots & K^{2p} \\ \vdots & \vdots & \ddots & \vdots \\ K^{l1} & K^{l2} & \cdots & K^{lp} \end{bmatrix}$$

A square exponential kernel function $K(\cdot)$ is employed as multi-output Gaussian process covariance function;

$$k(\mathbf{a}, \mathbf{a}^{'}; \boldsymbol{\theta}) = \alpha_1 \exp\left(-\frac{1}{2}\frac{\|\mathbf{a} - \mathbf{a}^{'}\|^2}{\alpha_2}\right) \tag{3.1.13}$$

and its constituents are computed as $K_{ij} = k(\mathbf{a}, \mathbf{a}^{'}; \boldsymbol{\theta})$ here, $\mathbf{a}$ and $\mathbf{a}^{'}$ are multi-dimensional vector. Now, for 'n' queried points the posterior predictive distribution is given as $\tilde{\mathbf{y}}_i = \begin{bmatrix} \mathbf{Y}^{i1} & \mathbf{Y}^{i2} & \cdots & \mathbf{Y}^{iN} \end{bmatrix}^T$ at new design points $\tilde{\mathbf{x}}^i = \begin{bmatrix} \mathbf{X}^{i1} & \mathbf{X}^{i2} & \cdots & \mathbf{X}^{iM} \end{bmatrix}^T$. Considering noise-free case, the posterior predictions of multivariate Gaussian distribution at new input locations vector $\tilde{\boldsymbol{x}}^i$ can be illustrated as; $p(\tilde{\mathbf{y}}_i | \tilde{\mathbf{x}}^i; \mathbf{X}, \mathbf{Y}) \sim \mathcal{N}(\mu_{ij}, \Sigma_{ij})$ with predictive mean

μ_{ij} and variance Σ_{ij}.

$$\mu_{ij} = \mathbf{h}(\tilde{\mathbf{x}}^i; \boldsymbol{\theta}_j) + \mathbf{K}(\tilde{\mathbf{x}}^i; \mathbf{X}; \boldsymbol{\theta}_j) \cdot \mathbf{K}(\mathbf{X}, \mathbf{X}; \boldsymbol{\theta}_j)^{-1} \cdot (\mathbf{y}_j - \mathbf{h}(\mathbf{X}; \boldsymbol{\theta}_j)) \tag{3.1.14}$$

$$\Sigma_{ij} = \mathbf{K}(\tilde{\mathbf{x}}^i; \tilde{\mathbf{x}}^i, \boldsymbol{\theta}_j) + \mathbf{K}(\tilde{\mathbf{x}}^i; \mathbf{X}; \boldsymbol{\theta}_j) \cdot \mathbf{K}(\mathbf{X}, \mathbf{X}; \boldsymbol{\theta}_j)^{-1} \cdot \mathbf{K}(\mathbf{X}, \tilde{\mathbf{x}}^i; \boldsymbol{\theta}_j) \tag{3.1.15}$$

In equations 3.1.14 and 3.1.15 the i and j represents i^{th} queried point corresponding to j^{th} objective value.

3.2 Adaptive sampling with Bayesian optimization

Experiments are expensive and time-consuming the underlying challenge is obtaining optimal results in minimum iterations [63]. Optimization is key to the experimental design especially when multiple objectives are involved. In traditional experimental design modeling and optimization are distinct and conducted independently of each other. However, the challenging part is embedding the modeling into the optimization which can be overcome by adapting to the response surface with Bayesian optimization [15][21].

Bayesian optimization, a model-based global optimization approach plays a pivotal role when the objective function doesn't exist with a closed-form relationship (black-box) [44]. Bayesian optimization combines the Gaussian process and the acquisition function, the former serves as a probabilistic surrogate model which provides the information in the form of mean and variance associated with input location in the design space. And latter assists with the optimization that indicates the promising configuration in the input space that will result in optimal results based on information provided by the surrogate model [54]. As represented in Figure 4, Bayesian optimization starts with an initial belief about the objective function, expressed as a prior distribution. After observing data and considering this prior, we calculate a posterior distribution using Baye's rule i.e., updating our belief of the objective function. This updated belief guides the selection of the next sample by optimizing a utility function also known as the acquisition function in Bayesian optimization. The

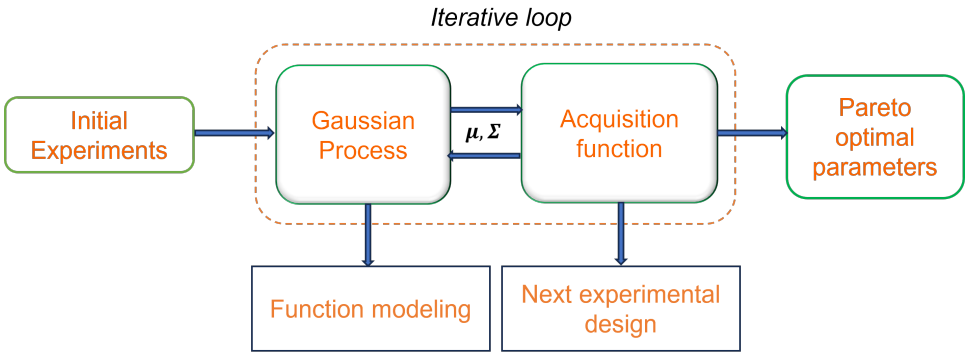

Figure 4: Sequential design with adaptive sampling

characteristics that make Bayesian optimization adaptive is balancing the trade-off between two measures exploration and exploitation to ascertain the position of the next sampling point:

- Exploration - Acquires data in regions that haven't been examined before, often guided by factors like proximity to existing data points. Additionally, with acquisition function sampling is done where the value of mean $\mu(\cdot)$ is high [15][21][63].

- Exploitation - Search regions where noteworthy patterns are detected, such as regions displaying random alterations or deviations from linearity. The next sampling point for the acquisition function will be the region with high variance $\sigma^2(\cdot)$ [15][21][63].

Several acquisition function in the Bayesian framework are as follow. If there exists a function $f : \mathcal{X} \to \mathbb{R}$ that we want to minimize in design space $X \subseteq \mathcal{X}$ then mathematical representation for global minimum is written as;

$$x^* = \arg\min_{x \in X} f(x) \tag{3.2.1}$$

In the subsequent sections, acquisition functions are defined for single and multiple BOs.

3.2.1 Single objective Bayesian optimization (SOBO)

Let's consider a minimization problem, $f^{'}$ denotes the best objective value (BOV) and f is the lowest value observed so far. Moreover, $\Phi(\cdot)$ and $\phi(\cdot)$ are CDF and PDF of Gaussian distribution, respectively. In section 3.2.1, we consider $x \in \mathbb{R}$, representing the design vector in a one-dimensional input space.

1) Probability of improvement (PI)

The initial acquisition function used under BO is the probability of improvement i.e., maximizing the probability of improvement over the best objective value observed so far and it is formulated as [15][63]:

$$u(x) = \begin{cases} 0 & \text{if } f(x) > f^{'} \\ 1 & \text{if } f(x) \leq f^{'} \end{cases}$$

$$PI(x) = P\left(f(x) \geq f^{'}\right) = \Phi\left(\frac{\mu(x) - f^{'}}{\sigma(x)}\right) \tag{3.2.2}$$

$u(x)$ represents utility function for x, it will be '0' (reward) if $f(x) > f^{'}$ or no reward otherwise. A point with a maximum of $PI(x)$ will be selected as an indication of improvement.

2) Expected Improvement (EI)

EI computes expectation $\mathbb{E}$ of improvement in relation to BOV ($f^{'}$) such that the utility function is maximized [15][63].

$$u(x) = \max(0, f^{'} - f(x)) \tag{3.2.3}$$

In the above utility function $u(x)$ we receive reward equivalent to improvement if $f(x)$ is less than $f^{'}$ and no reward otherwise.

$$EI(x) = \mathbb{E}\left[\max\left(0, f^{'} - f(x)\right)\right] = \left(f^{'} - \mu(x)\right)\Phi\left(\frac{f^{'} - \mu(x)}{\sigma(x)}\right) + \sigma(x)\phi\left(\frac{f^{'} - \mu(x)}{\sigma(x)}\right) \tag{3.2.4}$$

Initially, EI tends to focus on the initial best-observed value then converges to a global solution by balancing out two criteria exploration and exploitation by effectively utilizing the first two components of equation 3.2.4 i.e. increasing variance and reducing the mean.

3) Knowledge Gradient (KG)

In the knowledge gradient strategy, the goal is to maximize the expected utility value sequentially. However, it's important to note that this indicator is independent of the currently observed optimal value and tends to lean toward the distribution of the posterior mean [63][15].

$$KG(x) = \mathbb{E}_n[\max(\mu_{n+1}) - \max(\mu_n)] \tag{3.2.5}$$

The conditional expectation in relation to information gain after taking n measurement is given as $\mathbb{E}_n[\cdot] = \mathbb{E}[\cdot|X, Y]$

4) Entropy Search (ES)

Entropy search discovers the target region by reducing the uncertainty associated with optimal value x^* in input space by minimizing the entropy when new information is obtained at x and y represents output. Mathematically, it is denoted as;

$$H[p(x^*|D_n)] - \mathbb{E}_{p(y|D_n,x)}[H[p(x^*|D_n \cup \{(x, y)\})]] \tag{3.2.6}$$

In above equation, $H[p(x)] = -\int p(x)\log p(x)dx$ indicates measure of negative differential entropy associated with uncertainty x and $\mathbb{E}_p[\cdot]$ gives the expectation over probability

distribution (p) [15][63].

5) Upper confidence bound (UCB) and Lower confidence bound (LCB)

UCB and LCB form the criteria of bounds for maximization and minimization problems respectively.

$$UCB(x;\beta) = \mu(x) + \beta\sigma(x) \tag{3.2.7}$$

$$LCB(x;\beta) = \mu(x) - \beta\sigma(x) \tag{3.2.8}$$

where β is the hyper-parameter that balances the trade-off between exploitation and exploration captured by the mean μ and marginal standard deviation $\sigma(x) = \sqrt{\mathrm{Var}(x)}$ of Gaussian process. In the case of maximization, a higher value mean is preferred to satisfy criteria for exploitation, and a larger value of β leads to more exploration in unexplored regions within the design space as more reward is given to standard deviation [15][63].

3.2.2 Multi-objective Bayesian optimization (MOBO)

SOBO is no longer applicable for multi-objective schemes, as decision-makers are interested in all sets of possible solutions that can satisfy all objectives at once [55]. Some multi-objective problems involve conflicting and non-conflicting objectives therefore obtaining one solution could be a major challenge [42]. This limitation can be addressed by using the multi-objective optimization (MOO) method. In MOO, we obtain a Pareto front, which provides set solutions that are non-dominated i.e. one objective cannot be improved without affecting others hence the obtained solution is said as utopian [13]. Pareto sets are the collection of optimal solutions obtained in the variable space. The corresponding objective functions derived from these sets are referred to as the Pareto front [42][67].

MOO provides hierarchical association with sets of possible solutions. Following are some of the variations [66].

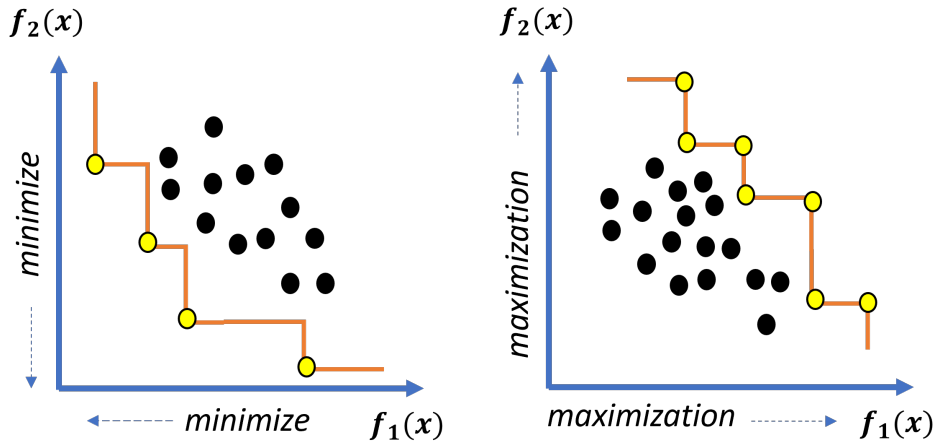

Figure 5: Illustration of Pareto sets (yellow dots), Pareto front (orange line), and non-dominated observations (black dots) in the objective space.

1) Dominance

Assuming case of the maximization problem, lets consider $\mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in \mathbb{R}^q$ sets of input vectors $\mathbf{y}^{(1)} = y(\mathbf{x}^{(1)})$ and $\mathbf{y}^{(2)} = y(\mathbf{x}^{(2)})$ are corresponding objectives, then we can state, $\mathbf{y}^{(1)}$ dominates $\mathbf{y}^{(2)}$, denoted as $\mathbf{y}^{(1)} \prec \mathbf{y}^{(2)}$ if $\forall i \in \{1, 2, \ldots, d\} : y_i(\mathbf{x}^{(1)}) \geq y_i(\mathbf{x}^{(2)})$ and $\exists j \in \{1, 2, \ldots, d\} : y_j(\mathbf{x}^{(1)}) > y_j(\mathbf{x}^{(2)})$ holds true [9].

2) Non-dominance

In the context of optimization and sequential exploration of the design space, decision-makers are more interested in non-dominated points [13]. Lets consider input vector $\mathbf{x} \in \mathbf{X}$, $\mathbf{X} \in \mathbb{R}^q$ and corresponding non-dominated output vector $\mathbf{y} \in nd(\mathbf{Y})$, $\mathbf{Y} = \{\mathbf{y}(\mathbf{x}) \,|\, \mathbf{x} \in \mathbf{X}\}$

$$nd(\mathbf{Y}) = \{\mathbf{y} \in \mathbf{Y} \,|\, \nexists \mathbf{z} \in \mathbf{Y} : \mathbf{z} \prec \mathbf{y}\} \tag{3.2.9}$$

3) Dominated space

Dominated subspace of Pareto front approximation $\mathcal{P}$ subset of $\mathbf{Y} \in \mathbb{R}^d$ is written as;

$$\text{dom}(\mathcal{P}) = \{\, \mathbf{y} \in \mathbb{R}^d \,|\, \exists \mathbf{p} \in \mathcal{P} \text{ with } \mathbf{p} \prec \mathbf{y} \,\} \tag{3.2.10}$$

4) Non-dominated space

Non-dominated set of pareto front with respect to $\mathbf{r}$ is denoted as ndom($\mathcal{P}$).

$$\text{ndom } (\mathcal{P}) := \{\mathbf{y} \in \mathbb{R}^d \,|\, \mathbf{y} \prec \mathbf{r} \text{ and } \nexists \mathbf{p} \in \mathcal{P} \text{ such that } \mathbf{p} \prec \mathbf{y}\} \tag{3.2.11}$$

where $\mathbf{r}$ is the reference point that is dominated by all the elements of the Pareto front in case of maximization problem and $\mathbf{r} \in \mathcal{P}$ and $\forall \mathbf{p} \in \mathcal{P} : \mathbf{p} \prec \mathbf{r}$.

5) Hypervolume Indicator (HV)

Pareto front, $\mathcal{P} = \{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)}\} \subset \mathbb{R}^d$ are evaluated by hypervolume indicator [67], it measures the area of dominated subspace enclosed by $\mathbf{r}$. HV is represented as;

$$\text{HV } (\mathcal{P}) = \lambda_d(\cup_{\mathbf{y} \in \mathcal{P}}[\mathbf{r}, \mathbf{y}]) \tag{3.2.12}$$

where λ_d denotes Lebesgue measures of hyper volume of set $\mathcal{P}$.

6) Hypervolume Improvement (HVI)

Hypervolume improvement is based on expected improvement, where we measure the volume of the Pareto front in the objective space [13]. The hypervolume indicator for the Pareto optimal solution set is given as;

$$\text{HVI } (\mathbf{y}, \mathcal{P}) = \text{ HV } (\mathcal{P} \cup \{\mathbf{y}\}) - \text{ HV } (\mathcal{P}) \tag{3.2.13}$$

7) Expected hypervolume Improvement (EHVI)

EHVI is defined as the integration of a finite collection of Pareto front over the hypervolume improvement [66].

$$EHVI(\mu, \sigma, \mathcal{P}, \mathbf{r}) = \int_{\mathbb{R}^d} \text{HVI } (\mathcal{P}, \mathbf{y}, \mathbf{r}) \cdot \xi_{\sigma,\mu}(\mathbf{y}) d\mathbf{y} \tag{3.2.14}$$

where, $\xi_{\sigma,\mu}$ represent the probability density function for multivariate normal distribution with mean μ and standard deviation σ.

8) Expected maximum minimum improvement (EMmI)

The multi-objective max-min improvement function $I_{\mathcal{M}}(\cdot)$, measures improvement of Pareto front relative to vectors of the design inputs $\mathbf{x} \in \mathcal{P}_{\mathcal{X}}$, the function is written as;

$$\mathbb{E}(I_{\mathcal{M}}(\mathbf{x})) = \mathbb{E}\Bigg[\max_{\mathbf{x}_i \in \mathcal{P}_{\mathcal{X}}} \min_{j=1,2,\cdots n} (\mathbf{y}_j(\mathbf{x}) - \mathbf{y}_j(\mathbf{x}_i)) \times \mathbb{1}\Bigg[\Bigg(\max_{\mathbf{x}_i \in \mathcal{P}_{\mathcal{X}}} \min_{j=1,2,\cdots n} (\mathbf{y}_j(\mathbf{x}) - \mathbf{y}_j(\mathbf{x}_i)) \Bigg) > 0 \Bigg] \Bigg] \tag{3.2.15}$$

From the above equation if $I_{\mathcal{M}} > 0$ then $\mathbf{x}$ is a dominant solution. Furthermore, it will be subpar to the solution in $\mathcal{P}_{\mathcal{X}}$ and if $I_{\mathcal{M}} = 0$ then the solution cannot be improved and is said to be weakly dominated. If $I_{\mathcal{M}} < 0$ then it results in a non-dominated solution for $\mathbf{x}$ which gives an improvement over the current set of solutions [1]. Now, '$\mathbb{1}$' is an indicator function with a binary operator.

From criteria for max-min improvement function and using the conditional distribution on fitness function of incumbent non-dominated points, $I_{\mathcal{M}}(\mathbf{x})$ [56].

The stopping criteria for EMmI, is either the budget assigned for experiments is attained, Pareto saturation i.e. identical parameters are generated during the iterative experiments, or the value of expected improvement over the iterated parameters of the criteria is negligible.

CHAPTER IV

SEQUENTIAL OPTIMIZATION: A GPR-MOBO APPROACH WITH EMmI CRITERIA

In Chapter 4, we implemented the proposed GPR-MOBO model using a conical nozzle geometry. As illustrated in figure 6, the sequential experimental framework is initiated with the selection of process parameters followed by initial experiments by two-level factorial design. As per section 4.2.1, fluid shear responses S_1, S_2, and S_3 are extracted at the X-Y-Z location. Then, initial training data are provided to the GPR-MOBO framework described in sections 4.3.1 and 4.3.2 respectively for the purpose of function modeling as well as determining the next experimental point. Finally, the experimental results with post-convergence analysis are evaluated in section 4.3.3.

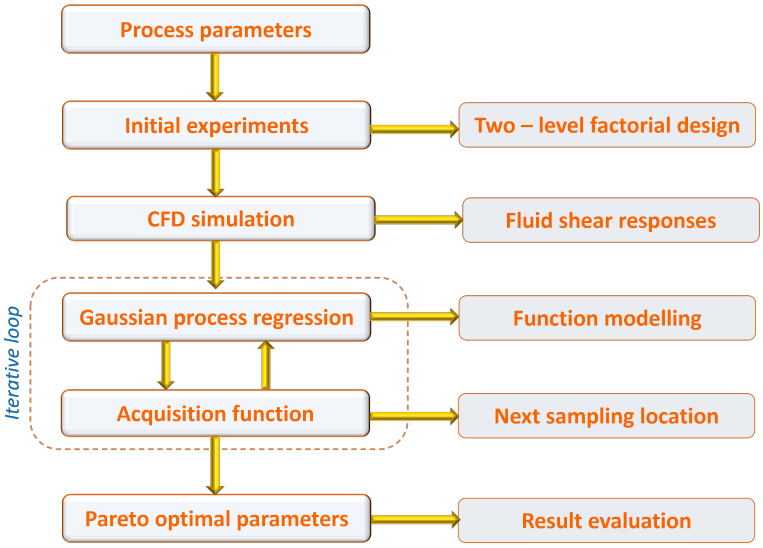

Figure 6: Sequential experimental framework

4.1 Experimental setup

4.1.1 Conical nozzle geometry

In our experimental setup, we conduct simulations using Ansys FLUENT to study extrusion-based bioprinting. Deposition in extrusion-based bio-printer is either droplet or continuous filament-based with either mechanical (piston or screw) or pneumatic deposition system. Bio-ink is carried out in a transient state through nozzle geometry also known as a syringe in fluid mechanics. Figure 6, illustrates conical nozzle geometry consisting of solid regions and a schematic diagram with dimensions (not drawn to scale). As shown in the figures,

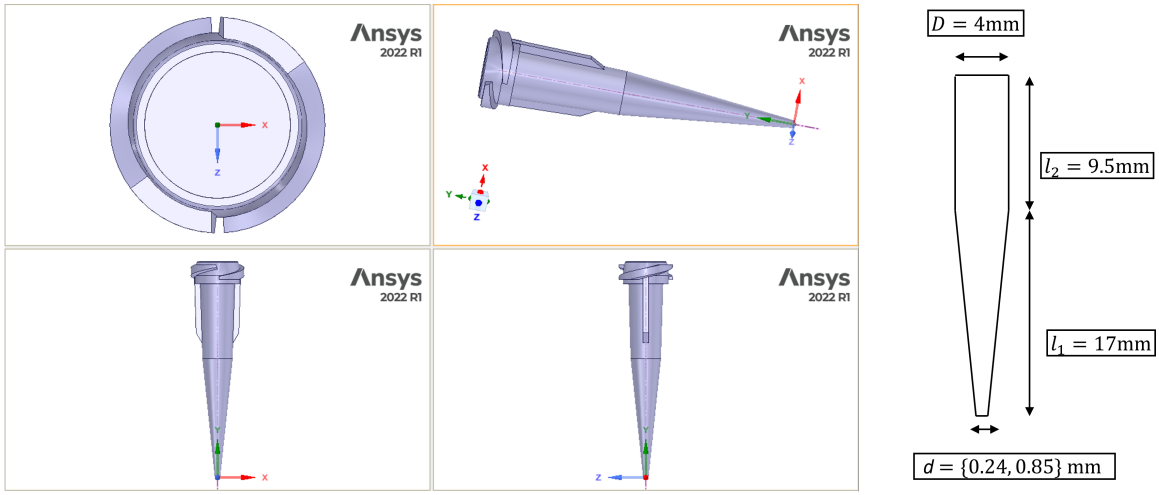

Figure 7: Conical nozzle geometry consisting of solid regions and a schematic diagram with dimensions

the convergent part of the nozzle is most fatal to cells during extrusion hence, we focus on examining the conical geometry in our simulations. Regarding the boundary conditions of the nozzle during simulations, the inlet diameter of geometry is fixed at $D = 4$mm, and the outlet diameter is varied as 18G and 25G in SI units $d \in \{0.24, 0.84\}$ mm. Additionally, l_1 is the converging length and l_2 is the straight section of the nozzle. In FLUENT simulation through a nozzle, the pressure boundary condition is computed through the inlet. Inlet pressure is set at 200000 Pa and outlet is atmospheric pressure at 101325 Pa [10]. Now, the inlet velocity is set at 0.028907 m/s. Finally, the material of the nozzle is plastic with properties such as

a density of 939.18 Kg/m^3 specific heat of 1844.7 $J/Kg-k$, and thermal conductivity of 0.26895 $W/m-k$. Throughout the simulation, the aforementioned parameters are constant except the outlet diameter is varied. Moreover, the details of mesh size are represented in Table 1. which is significant for capturing the geometry and behavior of the system under investigation. The element quality of the entire nozzle geometry is illustrated in Figure 8. Generally, the value ranges from 0 to 1 and the ideal element quality is the value closer to 1.

Parameter	**Value**
Nodes	6831
Elements	6052
Element size	1.6122 mm
Max size	3.2455 mm
Curvature size	0.016122 mm
Edge length	2.633 mm
Surface area	45.112 mm^2

Table 1: Details of mesh size

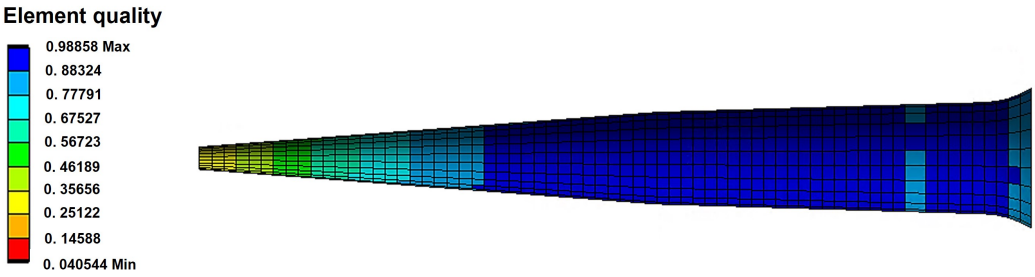

Figure 8: Element mesh quality

4.1.2 Bio-ink rheological properties

Bio-ink used for this experimental setup is composed of 60% nano–fibrillated cellulose (NFC) and 40% alginate, and the resulting water content in this solution is 97.5% [35]. Moreover, desirable bio-ink should have good mechanical stability, and shear thinning behavior in

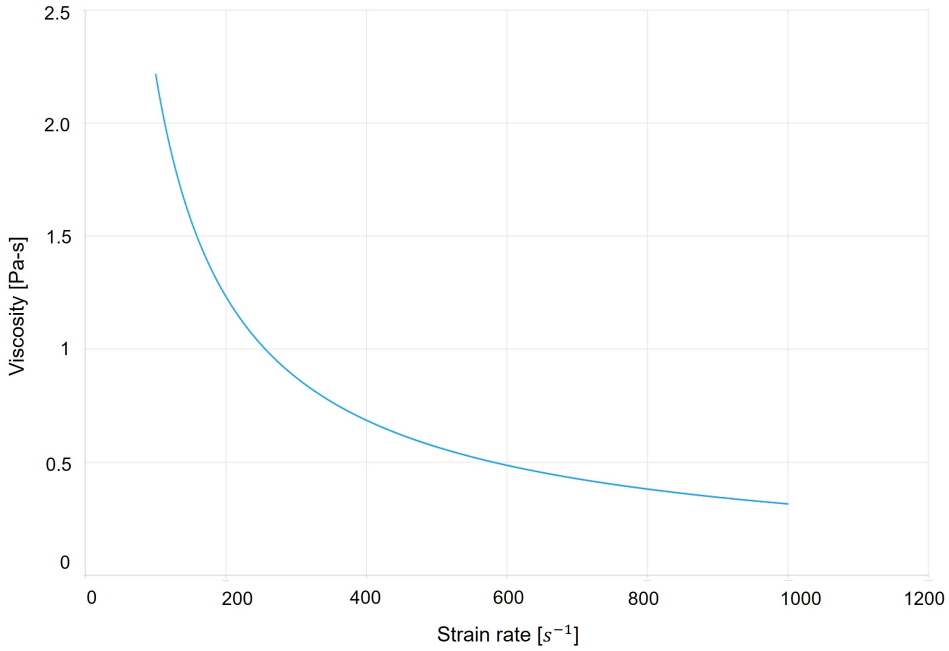

Figure 9: Bio-ink characteristics

addition to this it should create an environment for cells to thrive and survive.

$$\eta = \mathrm{K} \cdot \gamma^{(\mathrm{n}-1)} \tag{4.1.1}$$

The shear-thinning non-Newtonian behavior of the bio-ink is depicted in Figure 7, indicating that viscosity decreases as the shear rate increases. Furthermore, the laminar flow behavior of this bio-ink is modeled using a non-Newtonian power-law equation in Ansys FLUENT. In equation 4.1.1 represents a non-Newtonian power law model, η, K and n are viscosity, flow consistency, and flow behavior index respectively. As water has a viscosity of 0.00089 Pa.s and this bio-ink contains 97.5% water, therefore, the η_{min} is set at 0.00089 Pa-s and η_{max} as 62200 Pa.s. In the case of nozzle simulation in FLUENT, the bio-ink viscosity is varied from $\eta \in \{0.00089, 62200\}$ Pa.s and assuming bio-ink viscosity is not temperature dependent as a result, $(T_o = 0)$.The fluid material properties are also fixed at density 998.2 $Kg - m^{-3}$. Now throughout the experiment, $\mathrm{n} \in \{0.1, 1\}$ is changed. The upper bound for $\mathrm{n} \leq 1$ since, $\mathrm{n} > 1$ changes fluid behavior to shear thickening which is unfavorable, and $\mathrm{K} \in \{0.1, 200\}$

Pa.s are varied in the range to record the results for fluid shear responses [18][39][46].

4.2 Preliminary Experiments

As per ranges mentioned for process parameters in sections 4.1.1 and 4.1.2, initial experiments are designed with two-level factorial designs. Since, there are 5 factors and two levels $[-1, +1]$, we have a total run of 32 possible combinations resulting from 2^5 factorial design. As a result of these possible runs, we see from Table 2 in Appendix B, that for runs 17, 18, 19, 20, 25, 26, 27, 28: $\eta_{min} > \eta_{max}$ these conditions are infeasible therefore they are disregarded from our experiments. Furthermore, the 24 initial experiments are evaluated in Ansys to extract the internal fluid shear responses.

4.2.1 Design Methodology: Internal fluid shear stress

In this section, we present the simulation methodology in detail regarding extracting the information related to the internal fluid shear response of 18G and 24G nozzle. Further simulations are carried out considering the fluid region of the nozzle as shown in figure 10. Although the conical tip provides good print-ability with bio-ink of less viscosity the chal-

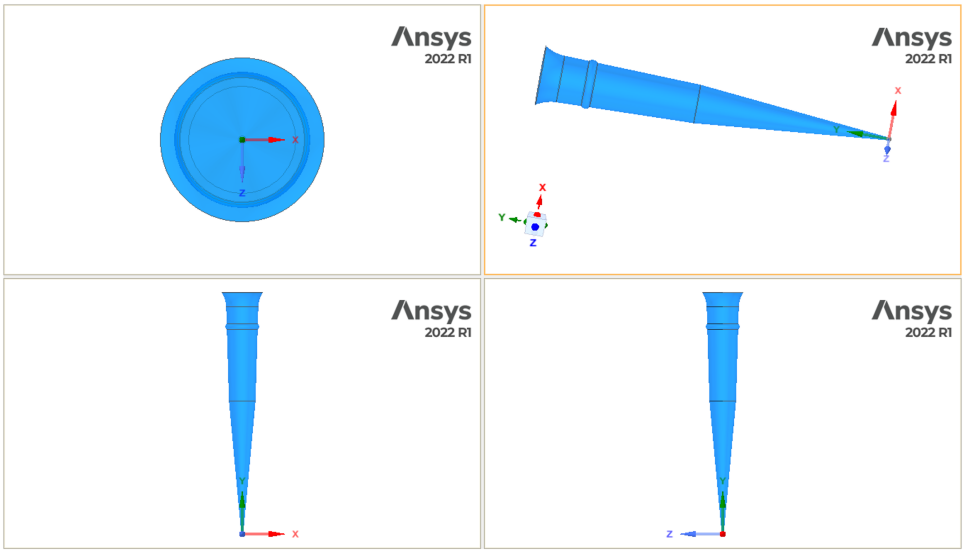

Figure 10: Conical nozzle geometry consisting of fluid region.

lenging part of this geometry is a tapered section that causes a change in shear distribution. Fluid mechanics theory for laminar flow suggests that for the straight section, there exists constant shear distribution. However, abrupt obstruction to the flow such as shoulder or necks against the pressurized fluid changes the shear pattern [18]. As shown in Figure 12 (A) we see significant wall shear is observed at the converging part of the nozzle. Our goal is to minimize the wall shear stress identified in this section therefore, we consider the conical section where the wall shear distribution changes randomly. As our model consists of a

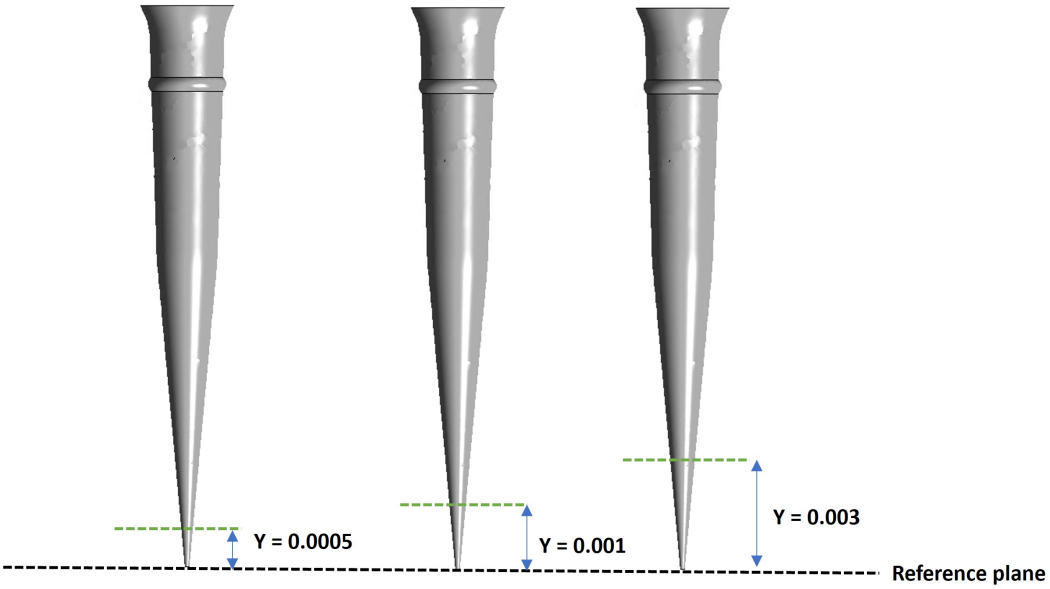

Figure 11: Vertical sections of fluid shear profile at $Y_1 = 0.0005$ mm, $Y_2 = 0.001$ mm, $Y_3 = 0.003$ mm

multi-objective framework we require three fluid shear responses where a shear transitioning pattern occurs. As a result, we deployed user-defined custom functions to model the relationship between wall shear and internal fluid shear stress. The relationship is given as in the case of laminar flow in FLUENT represented in equation 4.2.1. Here, τ, $\mu_{\text{molecular}}$, $\dot{\gamma}$ are shear stress, molecular viscosity, and strain rate respectively.

$$\tau \approx \mu_{\text{molecular}} \cdot \dot{\gamma} \tag{4.2.1}$$

Furthermore, the outlet tip of the nozzle is considered as the datum plane where three fluid shear responses are extracted in 'X' 'Y' and 'Z' direction. Here after the three shear responses are named 'S_1', 'S_2' and 'S_3' at locations ($X_1 = 0$, $Y_1 = 0.0005$ mm, $Z_1 = 0.00009$ mm), ($X_2 = 0$ mm, $Y_2 = 0.001$ mm, $Z_2 = 0.00015$ mm) and ($X_3 = 0$ mm, $Y_3 = 0.003$ mm, $Z_3 = 0.00025$ mm) respectively. Refer to Figures 11 and 12 (B) for a detailed illustration.

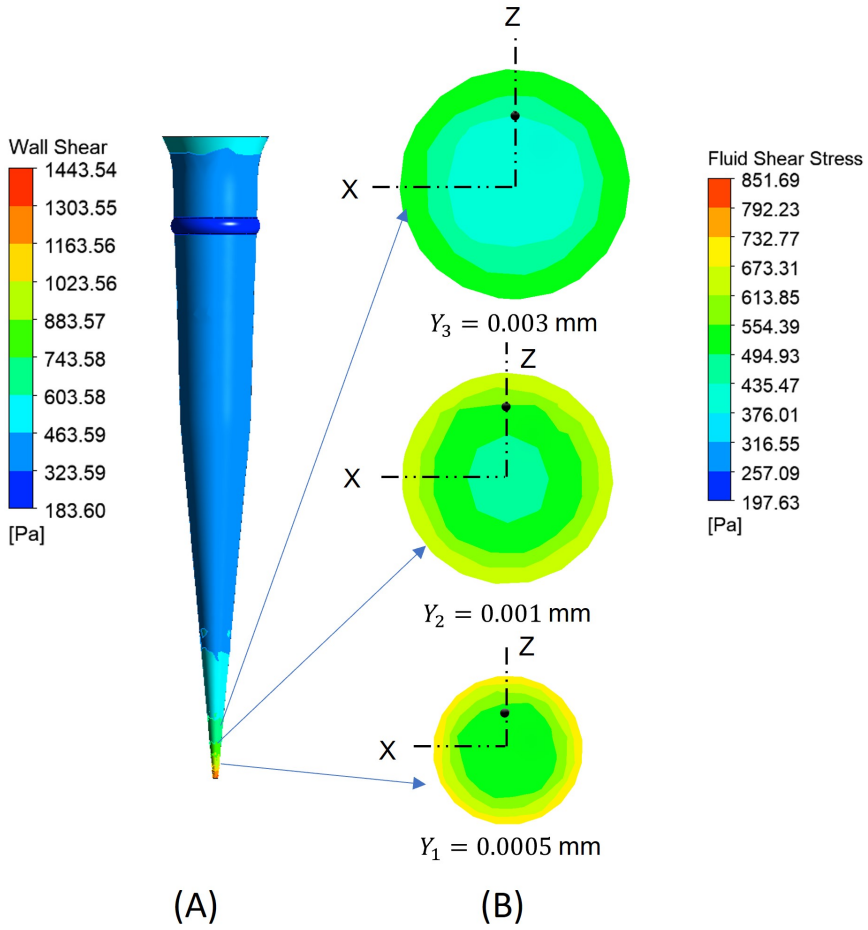

Figure 12: Wall and fluid shear profile of 25G conical nozzle at inputs $\eta_{max} = 62200$ Pa.s, $\eta_{min} = 0.00089$ Pa.s, n = 0.1 and K = 200 Pa.s [35].

4.3 Sequential experiments design of five design variables and three objectives

In this section, we briefly recall the GPR-MOBO scheme as per sections 3.1 and 3.2 respectively. Section 4.3.1 will describe how the process parameters are modeled with internal

fluid shear responses and section 4.3.2 will summarize the information related to the next sampling location. GPR-MOBO method and sequential experiments are implemented in R using the GPareto package [2][37].

4.3.1 Surrogate function modelling using Gaussian Process

In this sequential experimental design, our goal is to optimize the process parameters that will minimize internal fluid shear stress and eventually result in the minimization of wall shear. This will lead to print tissue construct with functional characteristics. In our design problem, we have three internal fluid shear responses denoted as S_1, S_2, and S_3. Let $\mathbf{y} \in \{S_1, S_2, S_3\}$ be the response vector consisting our objectives. Likewise, $\mathbf{x}$ denotes the design vector consisting of process parameters i.e. outlet diameter (mm), Viscosity (Pa.s), flow behavior index (n), flow consistency index (K). Furthermore, let $S_1 \sim f_1(\mathbf{x})$, $S_2 \sim f_2(\mathbf{x})$ and $S_3 \sim f_3(\mathbf{x})$. Thus, the optimization problem can be formulated as follows.

$$\mathbf{x}^* = \arg\min_{\mathbf{x}\in\mathcal{X}}(f_1(\mathbf{x}), f_2(\mathbf{x}), f_3(\mathbf{x})) \tag{4.3.1}$$

Afterward, we generated the initial experiments with factorial design refer to Table 2. Then, data is normalized within a scale of $[0, 1]$ to maintain the dimension uniformity. Now, this is a multi-objective problem without a closed form of relationship therefore we proceed with GP for the surrogate function modelling. The detailed description is introduced in section 3.1. Three objective functions are fitted individually with GP of zero mean and square exponential kernel function based on the initial 24 observations of experimental data from factorial design, refer to equation 3.1.4. The hyperparameters of the kernel function are optimized at each subsequent iteration with the maximum likelihood method to best fit the initial dataset. Furthermore, for each iteration mean and variance parameters are estimated.

4.3.2 Next sampling location by multi-objective Bayesian optimization

After defining the relationships between the process parameters and corresponding outputs we are inclined toward optimizing fluid shear responses to find the set of input parameters that minimizes our objectives. Therefore we are using MOBO using EMmI as acquisition criteria, refer to section 3.2 for a detailed description. Here, EMmI criteria are computed using Monte Carlo sampling since we have no closed-form relationship, and it samples from the posterior distribution of GP. Now, after initializing the 24 experiments surrogate model provides a predictive distribution in the form of mean and variance, and it serves as a prior belief for the set of functions that best fit the given data sets. Based on this information acquisition function provides the next sampling location and non-dominated points of internal fluid responses for the current iteration this represents the Pareto front. In addition, a value containing the possibility of dominating the current Pareto front is also provided by EMmI. Based on initial 24 experiments the incumbent Pareto front is obtained at runs 1 and 9, illustrated in Figure 13. The key point to highlight here, the nozzle outlet diameter is of standard sizes. Additionally, during the iterative experiments, EMmI recommends design inputs with fractional values. Therefore, to address this issue resultant design point is closely approximated to the nearest integer during the iteration to maintain diameter standardization [22][25][53]. This represents a non-dominated best objective value of internal fluid shear stress. This is nothing but average expected max-min improvement over current parameters, lower value or negligible updates in subsequent iterations denote the global Pareto optimal parameters are obtained and it acts as stopping criteria for the iterative loop. Meanwhile, we simultaneously maximize the acquisition function criteria until the model stops updating with the Pareto front. As acquisition functions require less computation time they are optimized with a genetic algorithm. Moreover, the population size is set to 160 because the number of inputs < 6 therefore $32 *$ number of inputs gives the required population size.

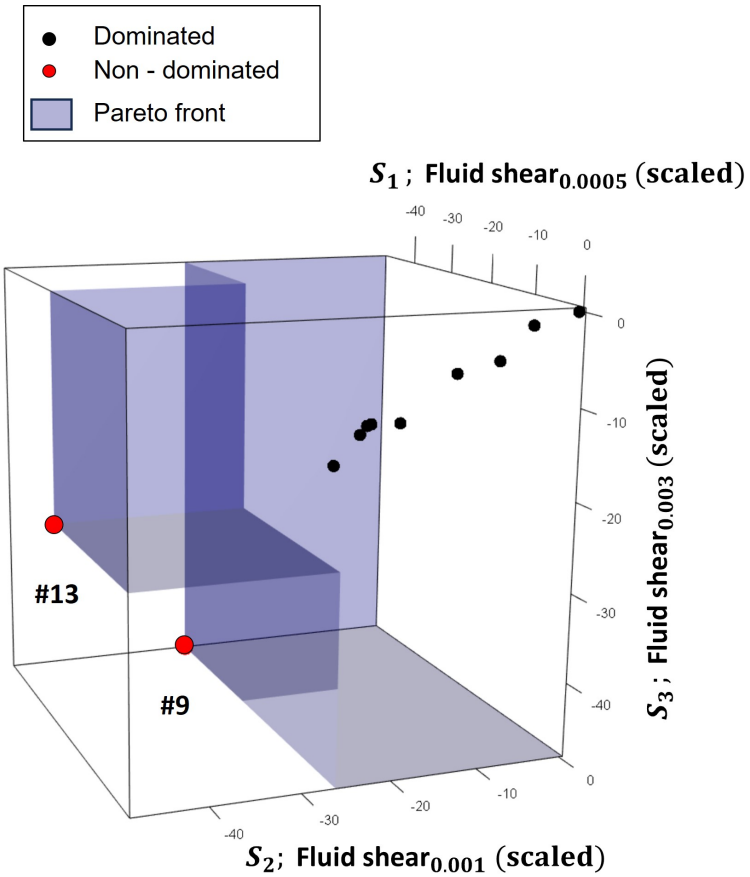

Figure 13: Incumbent Pareto front of preliminary experiments.

4.4 Results and Discussion

4.4.1 Resultant multi-objective Pareto front

We conclude that the GPR-MOBO approach converges after 21 iterations i.e. at the 45th run. Table 3 in Appendix C, depicts results obtained from sequential experiments with new recommended settings at each run along with the metric of expected improvement over the incumbent Pareto front. In addition to this Figure 14, illustrates the final multi-objective Pareto front after performing 21 sequential experiments. In this figure, the axes are scaled with logarithmic transformation ($S = \log(S) - 20$) which represents internal fluid shear responses at different 'Y' locations. The black points denote initial dominated designs. Furthermore, red and green points show non-dominated and Pareto-efficient solutions respectively with the Pareto front indicated as a blue front. The non-dominated solution stands for

best objective value and the shear thinning fluid (recommended) are set of solutions closer to the obtained Pareto front. Although some of the points evaluated are not in proximity to the Pareto front, most of them (indicated in green) are in the vicinity. It shows the efficacy of BO that does not waste its computational power in evaluating infeasible design points, this determines the BO as a sample-efficient method. The non-dominated points observed from this experiment are at the 1 and 9 runs. Additionally, the potential candidate solution that is closer to the Pareto front is at the 25th and 43rd runs. Specifically considering the experimental point of view the settings of run 1 and 9 are infeasible for further evaluations since the $\eta_{max} = \eta_{min}$ are equal which violates a primary condition of the non-Newtonian shear thinning fluid. As we know desirable characteristics of bio-ink indicate variation in viscosity in decreasing order as the shear rate increases as a result this suggest selecting the parameters with a constant setting is unfavorable. Therefore the recommended optimal parameter settings are the 25th and 43rd as we are neglecting non-dominated runs 1 and 9. The next potential candidate is the 43rd (non-Newtonian recommended shear thinning fluid) run after validating its internal fluid shear responses.

4.4.2 Pareto front saturation with EMmI strategy

Pareto saturation is the phenomenon where similar experimental input settings are generated in subsequent iterative runs which is observed in our approach. We infer the Pareto front has reached saturation considering the following observations from the illustration in Figure 15. The X-axis is the number of iterations while the corresponding Y-axis is the average expected improvement at each respective iteration. Firstly, the key point to highlight here is that significant improvements are not observed after the 12th iteration. Secondly, although the last candidate point added to the Pareto front is at the 43rd run i.e. 19th iteration which is identical to the previous design point at the 13th run. Thirdly, the parameters added at the 44th and 45th runs are exactly the same. Consequently, it indicates solid justification for Pareto saturation stating that further improvement is not possible. Finally, several

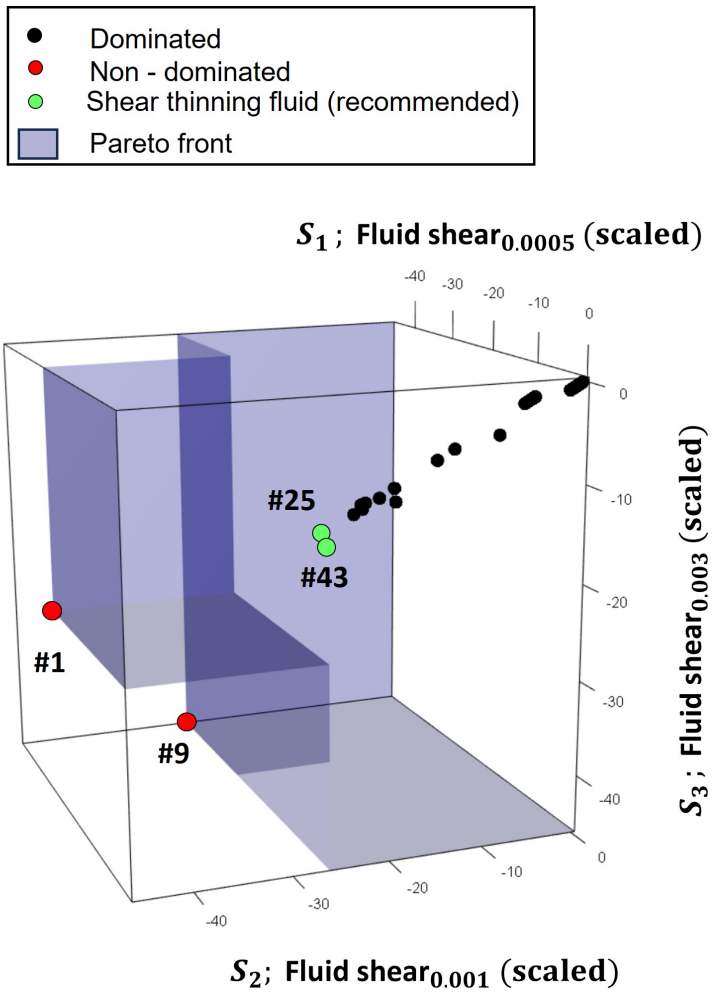

Figure 14: Multi-objective Pareto front of internal fluid responses with recommended settings.

other factors advocate our decision to stop performing subsequent iterations are satisfactory efforts made to discover new areas with the exploration and exploitation scheme of BO. It is illustrated in Figure 15, we can note that during the first 13 iterations, the graph exhibits fluctuations without noticeable trends. This signifies that the optimizer deploys both global and local searches during the process of optimization. Furthermore, after the 14th iteration average max-min improvement value criteria stop providing significant updates and converge at EMmI (X) 0.044 around the 21st iteration. This draws us to the conclusion of selecting runs 25 and 43 as prospective candidates for the optimal parameter exhibiting shear thinning behaviour.

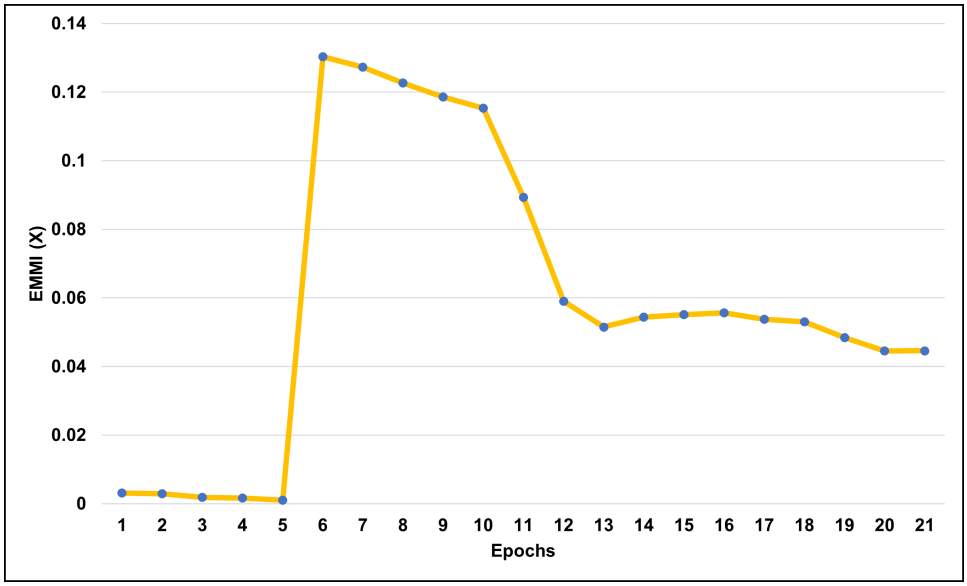

Figure 15: Average expected max - min improvement over recommended settings.

4.4.3 Recommended settings evaluation

The model recommends two settings at run 43 resulting process parameters; η_{max} = 62200 Pa.s, η_{min} = 0.00089 Pa.s, n = 0.1, K = 0.1 Pa.s and at run 25 resulting process parameters; η_{max} = 62156.60 Pa.s, η_{min} = 0.01276 Pa.s, n = 0.1027 and K = 0.256 respectively for 18G are evaluated in FLUENT. The optimal parameters that are obtained through the iterative GPR-MOBO framework can translated for real-world experiments. By scrutinizing the settings associated with each run some thoughtful results that can be drawn are as follows; high maximum viscosity suggests post-printing mechanical integrity of printed structure. Whereas, low viscosity indicates a smooth flow of bio-ink without clogging at the nozzle tip. Similarly, the value of n closer to 0.1 reveals shear thinning behaviour which is desirable for bio-ink extrusion. Moreover, the K values indicate the ability of the structure to maintain its form and shape. However, GPR-MOBO comes with one potential drawback, which is the requirement of domain expertise. Instances resulting from iterative experiments have value n = 1 and $\eta_{min} = \eta_{max}$ suggesting Newtonian fluid behavior. This result is an unfavorable condition for bio-ink due to Newtonian fluid behaviour and violates the non-Newtonian power law model. Researchers should evaluate such parameters considering feasibility in

laboratory experiments in terms of printer specification and requirement of tissue constructs. Additionally, an outlet nozzle diameter of 0.84 mm is preferred as a higher diameter suggests lower wall shear. The process parameters selected by the model balance out the internal trade-off to provide the lower value of resulting fluid shear. It is evident from Figure 16 that a significant reduction in internal fluid shear occurred at the sampling location in the X-Y-Z plane mentioned in section 4.2.1. Moreover, as we seek to minimize the internal fluid shear simultaneously the wall shear at the converging section of the nozzle is also minimized, the results of wall shear stress for run 25 of conical geometry are illustrated in Figure 16.

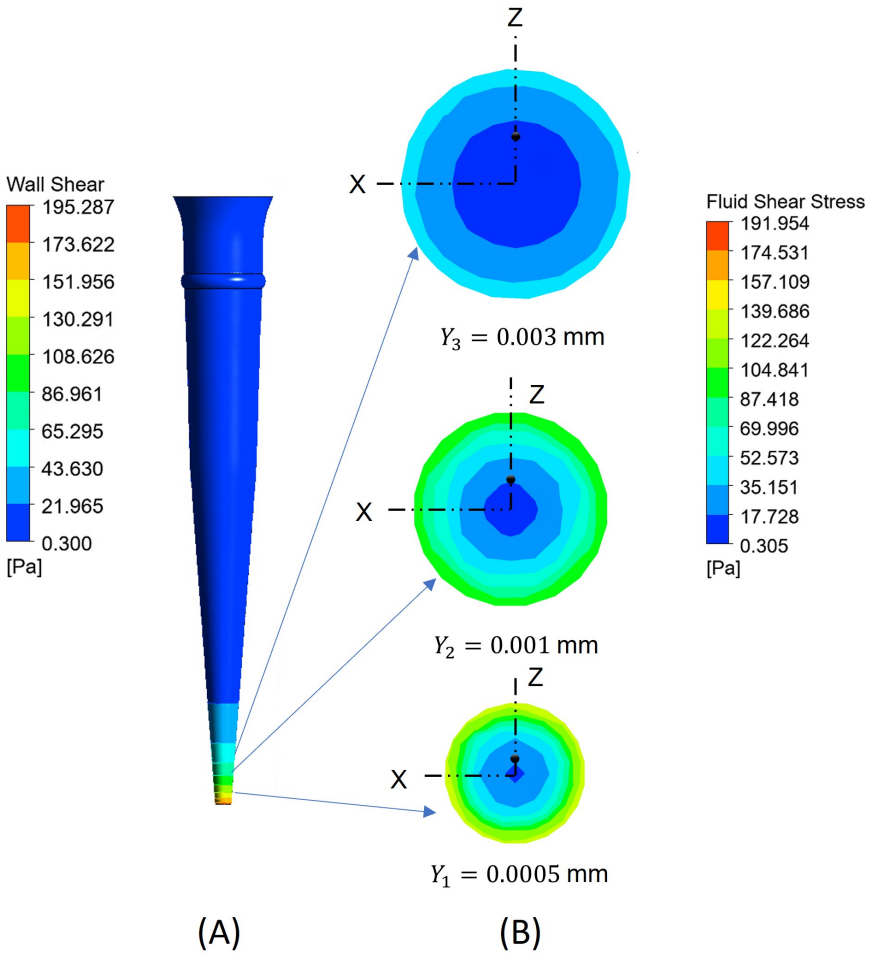

Figure 16: Wall shear and fluid shear of 18G conical nozzle at recommended inputs η_{max} = 62156.60 Pa.s, η_{min} = 0.0127 Pa.s, n = 0.1027 and K = 0.256 Pa.s.

CHAPTER V

CONCLUSION AND FUTURE WORK

5.1 Conclusion and Discussion

In this thesis, conical nozzle geometry of standard size 18G and 25G are examined for its applications in bioprinting by performing simulation analysis in FLUENT. Several process parameters such as print and rheological were considered to study the effect on cell functionality. The underlying assumption while conducting this research is higher the value of fluid shear stress more critical impact it will create on cell viability as a result printed cell construct will be less functional. Additionally, we minimize the maximum wall shear stress of the nozzle domain by minimizing the internal fluid shear responses through iterative experiments.

As there is non-existence of functional form relationship between the multi-dimensional variable and its corresponding objectives. To resolve this issue, we successfully implemented the GPR-MOBO framework. In which, GP served as a probabilistic surrogate model between process inputs and respective fluid shear responses at three locations in the X-Y-Z plane. In addition to this MOBO performed sequential optimization to determine the promising solution that satisfies our objectives. Moreover, convergence to Pareto optimal settings is achieved in 21 iterative experiments. After analyzing the convergence results we discovered that multiple objectives involved convey negligible trade-offs in addition to this they have high correlations at respective sampling locations i.e. $\text{corr}(Y_1, Y_2)$, $\text{corr}(Y_2, Y_3)$, $\text{corr}(Y_1, Y_3)$. For instance, let's consider our objectives S_1, S_2, and S_3, in this case minimizing S_1 and S_2 also minimizes S_3 and vice versa. Furthermore, as the Pareto non-dominant solution derived

from this method violates the condition of non-Newtonian shear thinning behavior of fluid we evaluated the next best solution that appears in proximity with the Pareto front. As they are the optimal parameter that satisfies pre-determined objectives along with required fluid behaviour. Overall, the GPR-MOBO model employed here is a concrete approach that guides the researchers in selecting the next experimental parameter. Nevertheless, the model requires further domain knowledge of parameter ranges employed to achieve resulting sampling points with precision and accuracy.

We conclude, that the resultant approach is novel in regards to bioprinting design application, for print parameter optimization of several nozzle geometries and bio-ink rheology. Specifically, in a domain where extrusion-based bio-printing is employed. Furthermore, it is evident that GPR-MOBO bypasses conventional trial and error optimization and reduces the time required for experimentation. In addition, it serves as an adaptive sampling method for creating an iterative framework and optimally selecting process inputs to minimize shear stress and maximize cell functionality. We expect that this study will be advantageous for the rapid development in the domain of bio-printing.

5.2 Future Work

In future work, it would be interesting to implement the GPR-MOBO framework in real-world experiments. Specifically, we propose collecting cell responses, such as protein secretion or the number of viable cells post-printing, to indicate cell functionality. Additionally, the investigation should explore the effects of more complex geometries, such as a co-axial nozzle that facilitates the simultaneous deposition of two different bio-inks. Within the domain of bioprinting, some design inputs are strictly integer values in our case outlet diameter of the nozzle. Due to the GP covariance function underlying assumption which takes inputs as continuous variables. In such cases, the kernel function should be user-defined to avoid sub-optimal results [17]. Interestingly, based on the results obtained, we have identified a potential opportunity to apply constraints to satisfy the non-Newtonian shear-thinning

behavior of bio-ink rheology. So, the potential input parameters could be sampling points that satisfy all constraints. However, violation of constraints will lead to assigning maximum values to objective functions. This will serve as a penalty term as a result optimizer will restrict itself from selecting such parameters. Some of these constraints are as follows:

- Minimum viscosity should be strictly less than maximum viscosity ($\eta_{min} < \eta_{max}$).
- Flow behavior index should be strictly less than 1 ($n < 1$).
- Flow consistency index K is constrained to be greater than or equal to K_{min} and less than or equal to K_{max} ($K_{min} \leq K \leq K_{max}$); where K_{min} and K_{max} are bounds on K.

This will modify our framework to "Domain-aware GPR-MOBO".

REFERENCES

[1] Dianne Carrol Bautista, *A sequential design for approximating the pareto front using the expected pareto improvement function*, The Ohio State University, 2009.

[2] Mickaël Binois and Victor Picheny, *Gpareto: An r package for gaussian-process-based multi-objective optimization and analysis*, Journal of Statistical Software **89** (2019), 1–30.

[3] George EP Box and Norman R Draper, *Response surfaces, mixtures, and ridge analyses*, John Wiley & Sons, 2007.

[4] George EP Box and Kenneth B Wilson, *On the experimental attainment of optimum conditions*, Breakthroughs in statistics: methodology and distribution, Springer, 1992, pp. 270–310.

[5] Rich Caruana, *Multitask learning*, Machine learning **28** (1997), 41–75.

[6] Rashik Chand, Beni Shimwa Muhire, and Sanjairaj Vijayavenkataraman, *Computational fluid dynamics assessment of the effect of bioprinting parameters in extrusion bioprinting*, 2022.

[7] Amit Choudhari, Aditya Rane, Shamir Talkar, Pawan Rayar, and Dhananjay Shukla, *Designing and prototyping for conservation and effective utilization of waste heat from air conditioner*, IOP Conference Series: Materials Science and Engineering, vol. 1104, IOP Publishing, 2021, p. 012007.

[8] Amit Choudhari, Shamir Talkar, Pavan Rayar, and Aditya Rane, *Design and manufacturing of compact and portable smart cnc machine*, Proceedings of International Conference on Intelligent Manufacturing and Automation: ICIMA 2020, Springer, 2020, pp. 201–210.

[9] Carlos A Coello Coello, *Evolutionary multi-objective optimization: basic concepts and some applications in pattern recognition*, Pattern Recognition: Third Mexican Conference, MCPR 2011, Cancun, Mexico, June 29-July 2, 2011. Proceedings 3, Springer, 2011, pp. 22–33.

[10] Varuna Dharmadasa, *Investigation of cell-viability in the bioprinting process*, 2016.

[11] Chuong B Do and Honglak Lee, *Gaussian processes*, Stanford University, Stanford, CA, accessed Dec **5** (2007), 2017.

[12] David Duvenaud, *Automatic model construction with gaussian processes*, 2014.

[13] Michael Emmerich, Kaifeng Yang, André Deutz, Hao Wang, and Carlos M Fonseca, *A multicriteria generalization of bayesian global optimization*, Advances in stochastic and deterministic global optimization (2016), 229–242.

[14] Ronald Aylmer Fisher, *Design of experiments*, British Medical Journal **1** (1936), no. 3923, 554.

[15] Peter I Frazier, *Bayesian optimization*, Recent advances in optimization and modeling of contemporary problems, Informs, 2018, pp. 255–278.

[16] Zhouquan Fu, Saman Naghieh, Cancan Xu, Chengjin Wang, Wei Sun, and Xiongbiao Chen, *Printability in extrusion bioprinting*, Biofabrication **13** (2021), no. 3, 033001.

[17] Eduardo C Garrido-Merchán and Daniel Hernández-Lobato, *Dealing with categorical and integer-valued variables in bayesian optimization with gaussian processes*, Neurocomputing **380** (2020), 20–35.

[18] Juan Carlos Gómez-Blanco, J Blas Pagador, Victor P Galván-Chacón, Luisa F Sánchez-Peralta, Manuel Matamoros, Alfonso Marcos, and Francisco M Sánchez-Margallo, *Computational simulation-based comparative analysis of standard 3d printing and conical nozzles for pneumatic and piston-driven bioprinting*, International Journal of Bioprinting **9** (2023), no. 4.

[19] Jochen Görtler, Rebecca Kehlbeck, and Oliver Deussen, *A visual exploration of gaussian processes*, Distill **4** (2019), no. 4, e17.

[20] Robert B Gramacy, *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*, CRC press, 2020.

[21] Stewart Greenhill, Santu Rana, Sunil Gupta, Pratibha Vellanki, and Svetha Venkatesh, *Bayesian optimization for adaptive experimental design: A review*, IEEE access **8** (2020), 13937–13948.

[22] Philipp Hennig and Christian J Schuler, *Entropy search for information-efficient global optimization.*, Journal of Machine Learning Research **13** (2012), no. 6.

[23] Charles Robert Hicks, *Fundamental concepts in the design of experiments*, Holt, Rinehart and Winston, New York,, 1964.

[24] WJ Hill, *Hunter. wg (1966)." a review of response surface methodology: A literature survey."*, Technometrics **8**, 571–590.

[25] Frank Hutter, Holger H Hoos, and Kevin Leyton-Brown, *Sequential model-based optimization for general algorithm configuration*, Learning and Intelligent Optimization: 5th International Conference, LION 5, Rome, Italy, January 17-21, 2011. Selected Papers 5, Springer, 2011, pp. 507–523.

[26] Pooja Jain, Himanshu Kathuria, and Nileshkumar Dubey, *Advances in 3d bioprinting of tissues/organs for regenerative medicine and in-vitro models*, Biomaterials **287** (2022), 121639.

[27] Hans-Michael Kaltenbach et al., *Statistical design and analysis of biological experiments*, Springer, 2021.

[28] Andre I Khuri, *A general overview of response surface methodology*, Biometrics & Biostatistics International Journal **5** (2017), no. 3, 87–93.

[29] Devara Venkata Krishna and Mamilla Ravi Sankar, *Persuasive factors on the bioink printability and cell viability in the extrusion-based 3d bioprinting for tissue regeneration applications*, 2023.

[30] Lucas Lemarié, Aravind Anandan, Emma Petiot, Christophe Marquette, and Edwin-Joffrey Courtial, *Rheology, simulation and data analysis toward bioprinting cell viability awareness*, Bioprinting **21** (2021), e00119.

[31] Haitao Liu, Jianfei Cai, and Yew-Soon Ong, *Remarks on multi-output gaussian process regression*, Knowledge-Based Systems **144** (2018), 102–121.

[32] Jia Liu, Jiafeng Ye, Fahim Momin, Xinyu Zhang, and Anyi Li, *Nonparametric bayesian framework for material and process optimization with nanocomposite fused filament fabrication*, Additive Manufacturing **54** (2022), 102765.

[33] Isabela Poley Magalhães, Patrícia Muniz de Oliveira, Janaína Dernowsek, Estevam Barbosa Las Casas, and Marina Spyer Las Casas, *Investigation of the effect of nozzle design on rheological bioprinting properties using computational fluid dynamics*, Matéria (Rio de Janeiro) **24** (2019).

[34] Ali Malekpour and Xiongbiao Chen, *Printability and cell viability in extrusion-based bioprinting from experimental, computational, and machine learning views*, Journal of Functional Biomaterials **13** (2022), no. 2, 40.

[35] Kajsa Markstedt, Athanasios Mantas, Ivan Tournier, Héctor Martínez Ávila, Daniel Hagg, and Paul Gatenholm, *3d bioprinting human chondrocytes with nanocellulose–alginate bioink for cartilage tissue engineering applications*, Biomacromolecules **16** (2015), no. 5, 1489–1496.

[36] R Mead, *A review of response surface methodology from a biometrics viewpoint*, J. Biometrics **31** (1975), 803–851.

[37] Walter R Mebane Jr and Jasjeet S Sekhon, *Genetic optimization using derivatives: the rgenoud package for r*, Journal of Statistical Software **42** (2011), 1–26.

[38] Tarun Shyam Mohan, Pallab Datta, Sepehr Nesaei, Veli Ozbolat, and Ibrahim T Ozbolat, *3d coaxial bioprinting: process mechanisms, bioinks and applications*, Progress in Biomedical Engineering **4** (2022), no. 2, 022003.

[39] Michael Müller, Ece Öztürk, Øystein Arlov, Paul Gatenholm, and Marcy Zenobi-Wong, *Alginate sulfate–nanocellulose bioinks for cartilage bioprinting applications*, Annals of biomedical engineering **45** (2017), 210–223.

[40] Raymond H Myers, Andre I Khuri, and Walter H Carter, *Response surface methodology: 1966–l988*, Technometrics **31** (1989), no. 2, 137–157.

[41] Raymond H. Myers and Douglas C. Montgomery, *Response surface methodology: Process and product in optimization using designed experiments*, 1st ed., John Wiley & Sons, Inc., USA, 1995.

[42] P. Ngatchou, A. Zarei, and A. El-Sharkawi, *Pareto multi objective optimization*, Proceedings of the 13th International Conference on, Intelligent Systems Application to Power Systems, 2005, pp. 84–91.

[43] Simon Olofsson, Mohammad Mehrian, Roberto Calandra, Liesbet Geris, Marc Peter Deisenroth, and Ruth Misener, *Bayesian multiobjective optimisation with mixed analytical and black-box functions: Application to tissue engineering*, IEEE Transactions on Biomedical Engineering **66** (2018), no. 3, 727–739.

[44] Nicolai Palm, Markus Landerer, and Herbert Palm, *Gaussian process regression based multi-objective bayesian optimization for power system design*, Sustainability **14** (2022), no. 19, 12777.

[45] Seongeon Park, Jonggeol Na, Minjun Kim, and Jong Min Lee, *Multi-objective bayesian optimization of chemical reactor design using computational fluid dynamics*, Computers & Chemical Engineering **119** (2018), 25–37.

[46] Axel Pössl, David Hartzke, Thomas M Schmidts, Frank E Runkel, and Peggy Schlupp, *A targeted rheological bioink development guideline and its systematic correlation with printing behavior*, Biofabrication **13** (2021), no. 3, 035021.

[47] S Raissi and R-Eslami Farsani, *Statistical process optimization through multi-response surface methodology*, International Journal of Mathematical and Computational Sciences **3** (2009), no. 3, 197–201.

[48] Carl Edward Rasmussen and Christopher K. I. Williams, *Gaussian processes for machine learning (adaptive computation and machine learning)*, The MIT Press, 2005.

[49] Kalani Ruberu, Manisha Senadeera, Santu Rana, Sunil Gupta, Johnson Chung, Zhilian Yue, Svetha Venkatesh, and Gordon Wallace, *Coupling machine learning with 3d bioprinting to fast track optimisation of extrusion printing*, Applied Materials Today **22** (2021), 100914.

[50] Thomas J Santner, Brian J Williams, William I Notz, Thomas J Santner, Brian J Williams, and William I Notz, *Space-filling designs for computer experiments*, The design and analysis of computer experiments (2018), 145–200.

[51] Keka Sinha, Shamik Chowdhury, Papita Das Saha, and Siddhartha Datta, *Modeling of microwave-assisted extraction of natural dye from seeds of bixa orellana (annatto) using response surface methodology (rsm) and artificial neural network (ann)*, Industrial Crops and Products **41** (2013), 165–171.

[52] Keka Sinha, Papita Das Saha, and Siddhartha Datta, *Response surface optimization and artificial neural network modeling of microwave assisted natural dye extraction from pomegranate rind*, Industrial crops and products **37** (2012), no. 1, 408–414.

[53] Jasper Snoek, Hugo Larochelle, and Ryan P Adams, *Practical bayesian optimization of machine learning algorithms*, Advances in neural information processing systems **25** (2012).

[54] Alexandros Solomou, Guang Zhao, Shahin Boluki, Jobin K Joy, Xiaoning Qian, Ibrahim Karaman, Raymundo Arróyave, and Dimitris C Lagoudas, *Multi-objective bayesian materials discovery: Application on the discovery of precipitation strengthened niti shape memory alloys through micromechanical modeling*, Materials & Design **160** (2018), 810–827.

[55] Wolfram Stadler, *Multicriteria optimization in engineering and in the sciences*, vol. 37, Springer Science & Business Media, 1988.

[56] Joshua Svenson and Thomas Santner, *Multiobjective optimization of expensive-to-evaluate deterministic computer simulator models*, Computational Statistics & Data Analysis **94** (2016), 250–264.

[57] Gen Taguchi et al., *Taguchi methods: design of experiments*, Amer Supplier Inst, 1993.

[58] Rania Taymour, David Kilian, Tilman Ahlfeld, Michael Gelinsky, and Anja Lode, *3d bioprinting of hepatocytes: Core–shell structured co-cultures with fibroblasts for enhanced functionality*, Scientific Reports **11** (2021), no. 1, 5130.

[59] Shuyu Tian, Rory Stevens, Bridget T McInnes, and Nastassja A Lewinski, *Machine assisted experimentation of extrusion-based bioprinting systems*, Micromachines **12** (2021), no. 7, 780.

[60] Jay M Ver Hoef and Ronald Paul Barry, *Constructing and fitting models for cokriging and multivariable spatial prediction*, Journal of Statistical Planning and Inference **69** (1998), no. 2, 275–294.

[61] Sanjairaj Vijayavenkataraman, Wei-Cheng Yan, Wen Feng Lu, Chi-Hwa Wang, and Jerry Ying Hsi Fuh, *3d bioprinting of tissues and organs for regenerative medicine*, Advanced drug delivery reviews **132** (2018), 296–332.

[62] Jie Wang, *An intuitive tutorial to gaussian processes regression*, 2022.

[63] Xilu Wang, Yaochu Jin, Sebastian Schmitt, and Markus Olhofer, *Recent advances in bayesian optimization*, ACM Computing Surveys **55** (2023), no. 13s, 1–36.

[64] He-Qi Xu, Jia-Chen Liu, Zheng-Yi Zhang, and Chang-Xue Xu, *A review on cell damage, viability, and functionality during 3d bioprinting*, Military Medical Research **9** (2022), no. 1, 70.

[65] Liang Yan, Xiaojun Duan, Bowen Liu, and Jin Xu, *Gaussian processes and polynomial chaos expansion for regression problem: linkage via the rkhs and comparison via the kl divergence*, Entropy **20** (2018), no. 3, 191.

[66] Kaifeng Yang, Michael Emmerich, André Deutz, and Thomas Bäck, *Efficient computation of expected hypervolume improvement using box decomposition algorithms*, Journal of Global Optimization **75** (2019), 3–34.

[67] Eckart Zitzler and Lothar Thiele, *Multiobjective evolutionary algorithms: a comparative case study and the strength pareto approach*, IEEE transactions on Evolutionary Computation **3** (1999), no. 4, 257–271.

APPENDICES

Appendix A. Multivariate Gaussian

$$X \sim \mathcal{N}(\mu_X, \Sigma_{XX}) \tag{A.1}$$

$$Y \sim \mathcal{N}(\mu_Y, \Sigma_{YY}) \tag{A.2}$$

In equation (6), μ and Σ represent the mean vector and covariance matrix respectively of random variables X and Y.

$$P_{X,Y} = \begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma) = \mathcal{N}\left(\begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix}, \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix} \right) \tag{A.3}$$

And, their marginalized density can be given as;

$$p_X(x) = \int_y p_{X,Y}(x, y)\, dy \tag{A.4}$$

$$p_Y(y) = \int_x p_{X,Y}(x, y)\, dx \tag{A.5}$$

The equation above can be understood as follows: If we want to find the probability density of the random variable X taking on a specific value x, we need to consider all the observations within our data set that account for the possible values of the related variable Y and vice versa. Similarly, conditioning can be defined as,

$$X \,|\, Y \sim \mathcal{N}\left(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - \mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}\right) \tag{A.6}$$

$$Y \mid X \sim \mathcal{N}\left(\mu_Y + \Sigma_{YX}\Sigma_{XX}^{-1}(X - \mu_X), \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right) \tag{A.7}$$

Appendix B. Preliminary initial experiments with two-level factorial design

Run	d (mm)	η_{max} (Pa.s)	η_{min} (Pa.s)	n	K (Pa.s)
1	0.25	0.00089	0.00089	0.1	0.1
2	0.25	0.00089	0.00089	0.1	200
3	0.25	0.00089	0.00089	1	0.1
4	0.25	0.00089	0.00089	1	200
5	0.25	62200	0.00089	0.1	0.1
6	0.25	62200	0.00089	0.1	200
7	0.25	62200	0.00089	1	0.1
8	0.25	62200	0.00089	1	200
9	0.84	0.00089	0.00089	0.1	0.1
10	0.84	0.00089	0.00089	0.1	200
11	0.84	0.00089	0.00089	1	0.1
12	0.84	0.00089	0.00089	1	200
13	0.84	62200	0.00089	0.1	0.1
14	0.84	62200	0.00089	0.1	200
15	0.84	62200	0.00089	1	0.1
16	0.84	62200	0.00089	1	200
17	0.25	0.00089	62200	0.1	0.1
18	0.25	0.00089	62200	0.1	200
19	0.25	0.00089	62200	1	0.1
20	0.25	0.00089	62200	1	200
21	0.25	62200	62200	0.1	0.1
22	0.25	62200	62200	0.1	200
23	0.25	62200	62200	1	0.1
24	0.25	62200	62200	1	200
25	0.84	0.00089	62200	0.1	0.1
26	0.84	0.00089	62200	0.1	200
27	0.84	0.00089	62200	1	0.1
28	0.84	0.00089	62200	1	200
29	0.84	62200	62200	0.1	0.1
30	0.84	62200	62200	0.1	200
31	0.84	62200	62200	1	0.1
32	0.84	62200	62200	1	200

Table 2: Initial experiments with factorial design

Appendix C. Results from sequential experiments

Run	d (mm)	η_{max} (Pa.s)	η_{min} (Pa.s)	n	K (Pa.s)	EMMI (X)
25	0.84	62156.61	0.012	0.10	0.17	0.0030
26	0.84	61861.89	1.84	0.99	199.54	0.0029
27	0.25	62200	54303.49	0.1	0.1	0.0018
28	0.25	62200	47240.63	0.1	0.1	0.0016
29	0.25	62200	40442.71	0.1	0.1	0.0010
30	0.25	62200	33776.70	0.1	0.1	0.1303
31	0.25	62200	58354.43	0.1	0.1	0.1272
32	0.25	62200	26882	0.1	0.1	0.1226
33	0.84	62200	30028.32	0.1	0.1	0.1185
34	0.25	62200	30031.71	0.1	200	0.1153
35	0.84	62200	43782.25	0.1	0.1	0.0893
36	0.84	62200	51098.43	0.1	200	0.0589
37	0.84	62200	37337.08	0.1	200	0.0514
38	0.84	62200	56407.73	1	200	0.0544
39	0.25	62200	43635.26	0.1	200	0.0551
40	0.84	62200	50553.50	1	0.1	0.0558
41	0.84	169.16	0.6829	0.10	0.37	0.0537
42	0.84	195.21	1.0858	0.99	2.17	0.0529
43	0.84	62200	0.00089	0.1	0.1	0.0483
44	0.84	62200	0.00089	1	0.1	0.0445
45	0.84	62200	0.00089	1	0.1	0.0445

Table 3: Results from iterative experiments with GPR-MOBO

VITA

ADITYA RANE

Candidate for the Degree of

Master of Science

Thesis: SURROGATE-ASSISTED MULTI-OBJECTIVE BAYESIAN OPTIMIZATION OF BIO-PRINTING PROCESS PARAMETERS

Major Field: Industrial Engineering and Management

Biographical:

Education:

Completed the requirements for the Master of Science in Industrial Engineering and Management at Oklahoma State University, Stillwater , Oklahoma in December 2023.

Completed the requirements for the Bachelor of Science in Mechanical Engineering at Mumbai University, Mumbai, Maharashtra in May 2018.

Experience:

- Graduate TA/RA School of Industrial Engineering and management (August 2022 - December 2023)
- Radcorps Fire and Safety Pvt Ltd - Junior Engineer (August 2020 – April 2021).
- ImpelPro SCM Solutions Pvt Ltd - Supply Chain Management Trainee (August 2019 – July 2020).
- TSM Project Logistics Pvt ltd - Logistics Intern, Mumbai, India (June 2018 – December 2018).
- Brihanmumbai Electric Supply and Transport - Maintenance Engineer Trainee (June 2017 – July 2017).

Professional Membership: Society of Manufacturing Engineers (SME)