Antonio Espuña, Moisès Graells and Luis Puigjaner (Editors), Proceedings of the 27th European Symposium on Computer Aided Process Engineering – ESCAPE 27
October 1st - 5th, 2017, Barcelona, Spain

Bayesian Multi-Objective Optimisation of Neotissue Growth in a Perfusion Bioreactor Set-Up

Simon Olofsson[a], Mohammad Mehrian[b], Liesbet Geris[b], Roberto Calandra[c], Marc Peter Deisenroth[a] and Ruth Misener[a*]

[a]*Dept. of Computing, Imperial College London*
[b]*Dept. of Aerospace and Mechanical Engineering, Université de Liège*
[c]*Dept. of EECS, University of California, Berkeley*
** r.misener@imperial.ac.uk*

Abstract

We consider optimising bone neotissue growth in a 3D scaffold during dynamic perfusion bioreactor culture. The goal is to choose design variables by optimising two conflicting objectives: (i) maximising neotissue growth and (ii) minimising operating cost. Our contribution is a novel extension of Bayesian multi-objective optimisation to the case of one black-box (neotissue growth) and one analytical (operating cost) objective function, that helps determine, within a reasonable amount of time, what design variables best manage the trade-off between neotissue growth and operating cost. Our method is tested against and outperforms the most common approach in literature, genetic algorithms, and shows its important real-world applicability to problems that combine black-box models with easy-to-quantify objectives like cost.

Keywords: Bayesian optimisation, black-box optimisation, multi-objective optimisation, tissue engineering, bone neotissue engineering

1. Introduction

Bone tissue engineering (TE) develops methods for healing, improving or replacing damaged bone tissue. Modern developments open up the possibilities of personalised healthcare, which is expected to improve the quality and cost-effectiveness of healthcare, as it offers more personalised and targeted therapies (Fuentes-Garí et al., 2015). However, quality and cost-effectiveness are often competing objectives, and a trade-off between the two must be found to ensure the commercial viability of tissue-engineered products.

The application discussed in this paper is optimisation of design variables for bone neotissue growth on a bioreactor scaffold, where we have two conflicting objectives; maximising neotissue growth while minimising operating cost. The first objective function is an expensive-to-evaluate black box, modelled via a probabilistic surrogate model. The second objective function is analytical and cheap to evaluate. In Sec. 3 we demonstrate a method for finding solutions to this multi-objective optimisation (MOO) problem using a novel extension of Bayesian MOO methods, and in Sec. 4 we compare this method to a genetic algorithm (the most common approach in literature to solving MOO problems) and show that our method performs better for a collection of test problems as well as for the TE application.

2. Background

The following sections describe (i) our multi-objective problem, (ii) the basics of Bayesian optimisation, and (iii) different MOO strategies.

2.1. Objective Functions

In the process of neotissue growth in the scaffold during culture period, we are faced with a multi-objective problem where we try to maximise the percentage of the scaffold filled with neotissue (the filling) while minimising the amount of material (medium, growth factors, operator handling) used by the bioreactor process (the cost). The filling objective function is given by an *in silico* model (Guyot et al., 2015; Mehrian et al., 2017) that takes two design variables as inputs: (i) the refreshment period $r \in [12\text{h}, 96\text{h}]$ for bioreactor medium change, and (ii) the ratio $a \in [0\,\%, 100\,\%]$ of the medium changed every r hours.

We assume the cost is given by the total amount of bioreactor medium used for a 504-hour (21-day) experiment. The medium is changed a discrete number of times $\lfloor 504/r \rfloor$, and has an associated cost $c_\ell \approx \$0.50/\text{mL}$, with $V \approx 10\,\text{mL}$ of the medium in the system at any time. Thus, the total cost of one experiment is $c(r,a) = V \cdot c_\ell \cdot (1 + \lfloor 504/r \rfloor \cdot a)$. For simplicity, the constant $V \cdot c_\ell$ is ignored, and the cost simplifies to $c(r,a) = 1 + \lfloor 504/r \rfloor \cdot a \in [1, 43]$. We compute $\partial c/\partial a$ exactly and approximate $\partial c/\partial r$ by ignoring the floor operator.

2.2. Bayesian Optimisation and Gaussian Processes

We begin by looking at an optimisation method for a single objective function f. Bayesian optimisation (BO) is a global black-box optimisation method (Kushner, 1964), which is useful for life science applications (see e.g. Ulmasov et al. (2016)). Function evaluations are used to construct a Gaussian process (GP) model of the objective function. A GP is a collection of random variables, any finite subset of which are jointly Gaussian distributed (Rasmussen and Williams, 2006). We place a GP prior $GP(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ on the unknown function f, where $\mu(\cdot)$ and $k(\cdot,\cdot)$ is the mean and covariance function, respectively. The choice of $\mu(\cdot)$ and $k(\cdot,\cdot)$ fully specifies the GP prior.

The GP prior implies that the function values $[\mathbf{f}]_i = f(\mathbf{x}_i)$ at points $\{\mathbf{x}_i\}$ are jointly Gaussian distributed with the function value $f(\mathbf{x}_*)$ at a query point $\mathbf{x}_*$. Assume observations $\mathbf{y}$ at points $\{\mathbf{x}_i\}$ are made. GP regression uses Bayes' theorem to compute the posterior predictive distribution $\mathcal{N}(f(\mathbf{x}_*)|\mu_*, \sigma_*^2)$, where $\mu_* = \mu(\mathbf{x}_*) + \mathbf{k}_*^\top \mathbf{K}^{-1}(\mathbf{y} - \boldsymbol{\mu})$ and $\sigma_*^2 = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top \mathbf{K}^{-1} \mathbf{k}_*$, with, in turn, $[\boldsymbol{\mu}]_i = \mu(\mathbf{x}_i)$, $[\mathbf{K}]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ and $[\mathbf{k}_*]_i = k(\mathbf{x}_i, \mathbf{x}_*)$.

By maximising an acquisition function, which has a built-in trade-off between exploration and exploitation, we compute an optimal choice for the next function evaluation. BO often needs fewer function evaluations than many other optimisation methods. Possibly the most commonly used acquisition functions for single-objective BO is expected improvement (Mockus et al., 1978) $\mathbb{E}[y_{\max} - f(\mathbf{x}_*)] = \sigma_* (\tilde{\mu}\Phi(\tilde{\mu}) + \phi(\tilde{\mu}))$, where $\tilde{\mu} = (y_{\max} - \mu_*)/\sigma_*$, $y_{\max} = \max \mathbf{y}$, ϕ is the zero-mean, unit variance Gaussian probability density function and Φ is the corresponding cumulative distribution function.

2.3. Multi-Objective Optimisation

In MOO, the goal is to reach a compromise between multiple, conflicting objectives. Assume a D-dimensional variable space and n_f conflicting objective functions $f_i : \mathbb{R}^D \to \mathbb{R}$, where the goal is to find the input $\arg\min_{\mathbf{x}}\{f_i(\mathbf{x})\}$ that minimises the objective functions.

There are two distinctly different ways of finding an optimal trade-off between the objectives (Hwang and Masud, 1979): (i) scalarisation, where the trade-off is made *a priori* by choosing an aggregated, scalarised function f_s and weight coefficients $\{\omega_i\}$, e.g. the weighted sum $f_s(\mathbf{x}) = \sum \omega_i f_i(\mathbf{x})$, and (ii) computing the Pareto frontier (PF), where the trade-off is made *a posteriori* by selecting the input that yields the most satisfying Pareto-optimal output. Pareto-optimality is a state in which the value of one objective function cannot be improved without impairing the value of another.

Scalarisation is easy to employ, but introduces a new problem as the weight coefficients have to be chosen *a priori*, often with incomplete understanding of how the system performs. Additionally, an optimal solution for a specific set of weights yields no information about other possible optimal solutions for different sets of weights. Weighted sum scalarisation also suffers from the limitation that it can only converge on trade-offs lying on convex sections (where $\mathrm{d}^2 f_2/\mathrm{d}f_1^2 > 0$ for the case $n_f = 2$) of the PF (Messac et al., 2000), and if the entire PF is concave, only the extreme points at the boundaries can be found.

3. Method

We present a novel extension of the expected hyper-volume improvement (EHVI) acquisition function $\mathbb{E}_{HI}(\mathbf{x})$ (Emmerich et al., 2008). Our application has two conflicting objectives: maximising filling (i.e. minimising *negative* filling) and minimising cost.

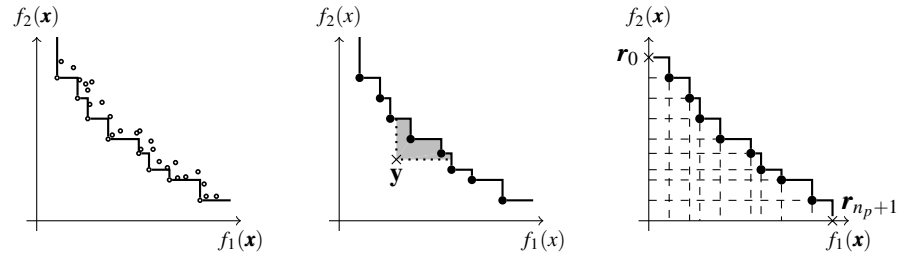
We assume that an approximated PF $\mathscr{P}_\alpha = \{\mathbf{r}_j\} = \{(r_{j,1}, r_{j,2})\}$, $j = 1,\ldots,n_p$, is given, with n_p being the number of non-dominated observations so far (see Fig. 1a). An observation $\mathbf{r}^{(1)} = (r_1^{(1)}, r_2^{(1)})$ dominates another observation $\mathbf{r}^{(2)} = (r_1^{(2)}, r_2^{(2)})$ if $\forall_k : r_k^{(1)} \leq r_k^{(2)}$ and $\exists_k : r_k^{(1)} < r_k^{(2)}$. Given $\mathscr{P}_\alpha$, Emmerich et al. (2008) define the EHVI as:

$$\mathbb{E}_{HI}(\mathbf{x}) = \int I_V(\mathbf{y}, \mathscr{P}_\alpha) p(\mathbf{y}|\mathbf{x}) \mathrm{d}\mathbf{y}, \tag{1}$$

where $I_V(\mathbf{y}, \mathscr{P}_\alpha) = \mathrm{Vol}(\mathscr{P}_\alpha \cup \mathbf{y}) - \mathrm{Vol}(\mathscr{P}_\alpha)$ is the area improvement given point $\mathbf{y}$ (see Fig. 1b), computed with respect to a user-defined reference point. Note that $I_V(\mathbf{y}, \mathscr{P}_\alpha) = 0$ if $\mathbf{y}$ is dominated by any point in $\mathscr{P}_\alpha$.

Cost $f_1(\mathbf{x})$ is deterministic, whereas filling $f_2(\mathbf{x}) \sim \mathscr{N}(\mu_2(\mathbf{x}), \sigma_2^2(\mathbf{x}))$ is modelled with a GP. Thus the probability of observing $\mathbf{y} = (y_1, y_2)$ is $p(\mathbf{y}|\mathbf{x}) = \delta(y_1 - f_1(\mathbf{x}))p(y_2|\mathbf{x})$, where $\delta(\cdot)$ is the Dirac delta function, which lets us derive a novel, closed-form expression for $\mathbb{E}_{HI}(\mathbf{x})$. For notational convenience, we will not write out the dependency on $\mathbf{x}$ of f_1, μ_2 and σ_2^2.

We assume that $\mathscr{P}_\alpha$ is sorted, i.e. $r_{j,1} < r_{j+1,1}$ and $r_{j,2} > r_{i+1,2}$ for all $j = 1,\ldots,n_p$, and introduce surrogate points $\mathbf{r}_0$ and $\mathbf{r}_{n_p+1}$ such that the available objective space is the rectangle with corners in $\mathbf{r}_0$ and $\mathbf{r}_{n_p+1}$. We divide the region below $\mathscr{P}_\alpha$ into rectangles C_{ij} defined by corners $(r_{i,1}, r_{j,2})$ and $(r_{i+1,1}, r_{j+1,2})$, as in Fig. 1c. The reference point for calculating $\mathrm{Vol}(\cdot)$ is $(r_{n_p+1,1}, r_{0,2})$. Now we utilise the fact that one objective function is



(a) (b) (c)

Figure 1: Illustrations of (a) an approximated PF $\mathscr{P}_\alpha$ (line) given observations (dots), (b) improvement (grey area) an observation $\mathbf{y}$ would yield to $\mathscr{P}_\alpha$, and (c) the region below $\mathscr{P}_\alpha$ divided into rectangles.

deterministic and that $p(\mathbf{y}|\mathbf{x}) = 0$ for all $y_1 \neq f_1$, and find the integer index $0 \leq h \leq n_p$ such that $r_{h,1} < f_1 \leq r_{h+1,1}$. Given this, Eq. (1) can be rewritten as:

$$\mathbb{E}_{HI}(\mathbf{x}) = \sum_{j=h}^{n_p} \int_{r_{j+1,1}}^{r_{j,1}} I_V((f_1, y_2), \mathscr{P}_\alpha) p(y_2|\mathbf{x}) \mathrm{d}y_2 \,. \tag{2}$$

Since $\mathbf{y} \in C_{ij}$ is in the region below $\mathscr{P}_\alpha$, it is not dominated but might instead dominate some points in $\mathscr{P}_\alpha$, and so $\mathrm{Vol}(\mathscr{P}_\alpha \cup \mathbf{y}) = \mathrm{Vol}\left(\{\mathbf{r}_0, \ldots, \mathbf{r}_i, \mathbf{y}, \mathbf{r}_{j+1}, \ldots, \mathbf{r}_{n_p+1}\}\right)$. Now define the area of the first n points in $\mathscr{P}_\alpha$ as $V_n = \sum_{i=1}^{n} (r_{n_p+1,1} - r_{i,1})(r_{i-1,2} - r_{i,2})$, where $V_0 = 0$ and $V_{n_p} = V_{n_p+1} = \mathrm{Vol}(\mathscr{P}_\alpha)$. We insert this into Eq. (2) and integrate out y_2, which yields the final expression:

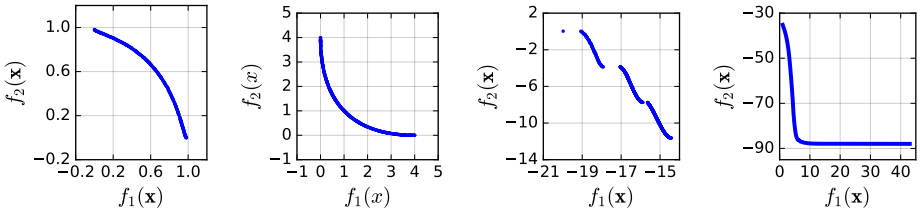
$$\begin{aligned} \mathbb{E}_{HI}(\mathbf{x}) = \sum_{j=h}^{n_p} \Big[\Big(& V_h - V_{j+1} + r_{h,2}(r_{n_p+1,1} - f_1) - r_{j+1,2}(r_{n_p+1,1} - r_{j+1,1}) \\ & + \mu_2(f_1 - r_{j+1,1}) \Big) \Phi - \sigma_2(f_1 - r_{j+1,1}) \phi \Big] \,, \end{aligned} \tag{3}$$

where $\Phi = \Phi\left(\frac{r_{j,2}-\mu_2}{\sigma_2}\right) - \Phi\left(\frac{r_{j+1,2}-\mu_2}{\sigma_2}\right)$ and $\phi = \phi\left(\frac{r_{j,2}-\mu_2}{\sigma_2}\right) - \phi\left(\frac{r_{j+1,2}-\mu_2}{\sigma_2}\right)$. The expression for the EHVI in Eq. (3) is easily differentiable with respect to f_1, μ_2 and σ_2. The objective functions are evaluated at $\arg\max_{\mathbf{x}} \mathbb{E}_{HI}(\mathbf{x})$ and $\mathscr{P}_\alpha$ recomputed.

4. Results

We compare the performance of our novel method to a different commonly used MOO method: genetic algorithms. Three different performance metrics are used: (i) the generational distance (GD) and (ii) maximum PF error (MPFE), which are measures of the average and maximum Euclidean distance, respectively, between points in $\mathscr{P}_\alpha$ to the true PF $\mathscr{P}_{\text{true}}$ (Van Veldhuizen, 1999), and (iii) the volume ratio (VR) $\mathrm{Vol}(\mathscr{P}_\alpha)/\mathrm{Vol}(\mathscr{P}_{\text{true}})$. Good performance results in low GD and MPFE, and a VR close to 1.

We select three different multi-objective test problems: (a) the Fonzeca and Fleming (1995) function for a two-dimensional input (concave PF), (b) the Schaffer (1984) two-objective function (convex PF), and (c) the Kursawe (1991) function (discontinuous PF). We also evaluate (d) the TE problem described in Sec. 2.1. Fig. 2 shows all of the PFs.



(a) Fonzeca and Fleming (b) Schaffer (c) Kursawe (d) TE

Figure 2: PFs for test problems (a), (b) and (c), and for the TE problem (d).

The Bayesian MOO method is given 10 random initial observations. The chosen genetic algorithm is the NGSA-III (Deb and Jain, 2014), which runs for 10 initial iterations. Both optimisation methods then run for 50 (additional) iterations. In Table 1, we compare the average performance (of 25 experiments) after 10, 25 and 50 iterations. We find that the Bayesian method consistently performs better.

5. Discussion

We have shown that the Bayesian MOO method performs better than the genetic algorithm for both the test problems and our TE application. This is unsurprising as the number of design variables and iterations are low, and because we are able to exploit the combination of black-box and analytical objective functions that occurs frequently in life sciences and process engineering.

The computational time required to choose the next query point has not been taken into account since our focus is on applications where evaluating the black-box function is significantly more expensive than optimising for the next query point. Whereas genetic algorithms can be preferable for applications with cheap function evaluations and higher number of design variables, Bayesian methods' maximal use of prior information and previous observations give them an advantage for problems such as the TE application discussed in this paper.

Table 1: Performance of the Bayesian method compared with the NSGA-III genetic algorithm, after 10, 25 and 50 iterations. Bold font denotes best average performance.

		GD		MPFE		VR	
	Iter.	EHVI	NSGA-III	EHVI	NSGA-III	EHVI	NSGA-III
Fonzeca and Fleming (1995)	10	**0.001**	0.1163	**0.1344**	0.7137	**0.8635**	0.185
	25	**0.0005**	0.0708	**0.0488**	0.5913	**0.9511**	0.3016
	50	**0.0004**	0.0393	**0.0244**	0.4356	**0.981**	0.4933
Schaffer (1984)	10	**0.0022**	0.0742	**0.5455**	1.4463	**0.9967**	0.984
	25	**0.0013**	0.031	**0.2992**	1.0601	**0.9989**	0.991
	50	**0.001**	0.0064	**0.1689**	0.5782	**0.9996**	0.9973
Kursawe (1991)	10	**0.0197**	3.4157	**1.4276**	9.1451	**0.9937**	0.5247
	25	**0.0068**	2.5024	**0.9333**	7.8127	**0.9979**	0.5834
	50	**0.0032**	1.3879	**0.6699**	6.4637	**0.9991**	0.6596
TE problem	10	**0.0764**	0.1321	**11.8454**	13.6394	**0.9877**	0.9737
	25	**0.0314**	0.0967	**10.871**	12.6811	**0.9932**	0.9802
	50	**0.0185**	0.067	**11.179**	12.2367	**0.9962**	0.9854

6. Conclusions

The novel Bayesian method outperforms competing MOO methods, for a diverse test set and for our tissue engineering application. The results show that the Bayesian method is highly applicable to real-world problems combining expensive black-box models with easy-to-quantify objectives like cost.

7. Acknowledgements

This work has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 675251, a Royal Academy of Engineering Research Fellowship to R.M., and a Google Faculty Research Award to M.P.D.

References

Deb, K., Jain, H., 2014. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints. IEEE Transactions on Evolutionary Computation 18 (4), 577–601.

Emmerich, M., Deutz, A., Klinkenberg, J.-W., 2008. The computation of the expected improvement in dominated hypervolume of pareto front approximations. Tech. rep., Leiden Institute for Advanced Computer Science.

Fonzeca, C. M., Fleming, P. J., 1995. An overview of evolutionary algorithms in multiobjective optimization. Evolutionary Computation 3 (1), 1–16.

Fuentes-Garí, M., Velliou, E., Misener, R., Pefani, E., Rende, M., Panoskaltsis, N., Mantalaris, A., Pistikopoulos, E. N., 2015. A systematic framework for the design, simulation and optimization of personalized healthcare: Making and healing blood. Computers & Chemical Engineering 81, 80–93.

Guyot, Y., Luyten, F., Schrooten, J., Papantoniou, I., Geris, L., 2015. A three-dimensional computational fluid dynamics model of shear stress distribution during neotissue growth in a perfusion bioreactor. Biotechnology and Bioengineering 112 (12), 2591–2600.

Hwang, C.-L., Masud, A. S. M., 1979. Multiple objective decision making – methods and applications: a state-of-the-art survey. Springer.

Kursawe, F., 1991. A variant of evolution strategies for vector optimization. In: Parallel Problem Solving from Nature. 1st Workshop, PPSN I, volume 496 of Lecture Notes in Computer Science. Springer-Verlag, pp. 193–197.

Kushner, H. J., 1964. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. Journal of Basic Engineering 86 (1), 97–106.

Mehrian, M., Guyot, Y., Papantoniou, I., Sonnaert, M., Geris, L., 2017. Model-based optimization of the medium refreshment regime during neotissue growth in a perfusion bioreactor. In: FOCAPO/CPC 2017.

Messac, A., Sundararaj, G. J., Tappeta, R. V., Renaud, J. E., 2000. Ability of objective functions to generate points on nonconvex pareto frontiers. AIAA Journal 38 (6), 1084–1091.

Mockus, J., Tieses, V., Zilinskas, A., 1978. The application of Bayesian methods for seeking the extremum. Towards Global Optimization 2, 117–129.

Rasmussen, C. E., Williams, C. K. I., 2006. Gaussian processes for machine learning. The MIT Press.

Schaffer, J. D., 1984. Some experiments in machine learning using vector evaluated genetic algorithms (artificial intelligence, optimization, adaptation, pattern recognition). Ph.D. thesis, Nashville, TN, USA.

Ulmasov, D., Baroukh, C., Chachuat, B., Deisenroth, M. P., Misener, R., 2016. Bayesian optimisation with dimension scheduling algorithm: application to biological systems. In: Kravanja, Z., Bogotaj, M. (Eds.), ESCAPE-26. Vol. 38 of Computer Aided Chemical Engineering. Elsevier, pp. 1051–1056.

Van Veldhuizen, D. A., 1999. Multiobjective evolutionary algorithms: classifications, analyses, and new innovations. Ph.D. thesis, Faculty of the Graduate School of Engineering of the Air Force Institute of Technology, Air University.