A Novel Derivative-Free Optimization Method based on Single Dimension Projection

Ishan Bajaj and M.M.Faruque Hasan*

Artie McFerrin Department of Chemical Engineering,
Texas A&M University, College Station, TX 77843

Abstract

An optimization method is proposed to solve a multi-dimensional black-box problem based on the projection onto a special 1-dimensional space. A univariate function on this space exists such that its optima corresponds to the optima of the original multi-dimensional problem. Based on sensitivity analysis, we provide useful properties of the function and geometric intuition on the conservation of the optima. A transformation matrix can be recursively applied to obtain this 1-dimensional function and transform back to the original n-dimensional space. A two-step iterative algorithm is also proposed to find the optima. The preliminary theoretical development shows promise to use this approach to effectively solve multi-dimensional black-box problems. The method is applied on 17 box-constrained test problems to demonstrate its effectiveness.

Keywords

Derivative-Free Optimization; Projection; Sensitivity Analysis, Black-box

Introduction

Many practical engineering problems are high-dimensional, computationally expensive, and may not have explicit functional forms/expressions. Examples include optimization of process operations described by high-fidelity models such as computational fluid dynamics, partial differential equations (Boukouvala et al., 2015) and flowsheet simulation (Caballero and Grossmann, 2008). The algebraic form of the objective function in these problems is unavailable and these can be referred as black-box problems (Nuchitprasittichai and Cremaschi (2011), Eason and Biegler (2016)). Although detailed process models have been useful for combining atomistic scale to process scale (Hasan et al., 2013), the computational costs of using gradient based solvers are high. As a result, there has been a growing interest in derivative-free optimization (DFO), both in theoretical advancements (Powell (2006), Conn et al. (2009), Wild and Shoemaker (2013)) and applications (Henao and Maravelias (2011), First et al. (2014)). It is assumed in DFO that evaluation of a black-box function is possible but expensive. One goal in solving these problems is to find the optimum using as few function evaluations as possible.

In spite of significant efforts in the past, the application of DFO methods has been limited to small dimensions because the problem complexity increases with increase in the number of variables (Rios and Sahinidis, 2013). Current strategies to handle high dimensional problems include reducing design space, screening significant variables, decomposing into smaller sub-problems and mapping (Shan and Wang, 2010). Mapping and decomposition have been reported as promising ways to tackle high-dimensional black-box problems. However, no previous work examined the possibility of mapping an optimization problem from the original high-dimensional space to a single dimension while preserving the optima.

This paper uses a projection of the original optimization problem to convert it into a univariate problem. By doing so, the problem size and complexity can be reduced simultaneously. We also show that the convexity,

*To whom all correspondence should be addressed *hasan@tamu.edu*, Tel.: 979-862-1449

concavity and twice-continuous differentiability properties of the original function hold true for the projected 1-dimensional function. Based on projection, we provide a novel algorithm to solve the following black-box problem:

$$\begin{aligned} \text{P1}: \quad & \min_{x} \quad f(x) \\ & s.t. \quad x_i \in [{x_i}^L, {x_i}^U] \quad \forall i \in \{1, \ldots, n\} \end{aligned} \tag{1}$$

where $x \in \mathbb{R}^n$, and $f(x) : \mathbb{R}^n \to \mathbb{R}$ is assumed to be twice-continuously differentiable black-box function.

The paper is organized as follows. We first provide the idea of projection and intuition on conservation of optima. Next, we present the properties of the projected function. The algorithm is then presented followed by computational results and concluding remarks.

Main Idea

We add an auxiliary variable t to P1 and re-define P1 as P2 as follows:

$$\begin{aligned} \text{P2}: \quad & \min_{x,t} \quad f(x) \\ & s.t. \quad \sum_{i=1}^{n} x_i = t \\ & \qquad x_i \in [{x_i}^L, {x_i}^U] \qquad \forall i \in \{1, \ldots, n\} \\ & \qquad t \in [\sum_{i=1}^{n} x_i^L, \sum_{i=1}^{n} x_i^U] \end{aligned} \tag{2}$$

Remark 1. Problem P2 is equivalent to P1 in a sense that both have same feasible, local and global optima sets because adding the auxiliary variable t does not alter the problem.

Therefore, we can also write P1 as P3 as follows:

$$\begin{aligned} \text{P3}: \quad & \min_{t} \quad F(t) \\ & s.t. \quad F(t) = \begin{bmatrix} \min\limits_{x} & f(x) \\ s.t. & \sum_{i=1}^{n} x_i = t \\ & x_i \in [{x_i}^L, {x_i}^U] \end{bmatrix} \\ & \qquad t \in [\sum_{i=1}^{n} x_i^L, \sum_{i=1}^{n} x_i^U] \end{aligned} \tag{3}$$

where, $t \in T \subset \mathbb{R}$ and *F(t)* is the 1-dimensional projected function.

Proposition 1. $\min\limits_{x} \; f(x) = \min\limits_{t} \; F(t)$

Proof. P3 $\Leftrightarrow$ P2, *P2* $\Leftrightarrow$ P1. Hence, P3 $\Leftrightarrow$ P1 $\square$

Geometrically, it can be interpreted as follows. If the samples (x, f) in the n-dimensional space is projected onto 1-dimension (t, f), multiple function values will be obtained corresponding to a particular t. Choosing the minimum of these values for each t results in a univariate function $(F(t))$ and optimizing this univariate function is equivalent to optimizing the original n-dimensional problem. This is illustrated in Figure 1 using 2-dimensional Branin function (Figure 1 (a)). Figure 1(b) is obtained by projecting all f values on t. The function has 3 minima and they can be observed in Figure 1(c) as well. If the exact points on $F(t)$ (shown by the red curve) can be obtained, the original problem is essentially reduced to univariate problem. Note that regardless of the problem dimension, a univariate function similar to that shown in Figure 1(c) always exists.

The problem P1 can be solved by decomposing into two sub-problems: solve the lower level problem for x for each t (P4) and then solve the upper level problem for t (P5). We explicitly define the two sub-problems as follows:

$$\begin{aligned} \text{P4}: \quad F(t) = \min_{x} \quad & f(x) \\ s.t. \quad & g_i^{(1)} := x_i^L - x_i \leq 0 \\ & g_i^{(2)} := x_i - x_i^U \leq 0 \\ & h_1 := \sum_{i=1}^{n} x_i = t \end{aligned} \tag{4}$$

$$\begin{aligned} \text{P5}: \quad & \min_{t} \quad F(t) \\ & s.t. \quad t \in [t^L, t^U] \end{aligned} \tag{5}$$

Solving P5 involves optimizing a trivial 1-dimensional black-box function. Since obtaining analytical expression of $F(t)$ is not possible for black-box problems, the idea is to solve P4 at each sampled t. In this work, we solve the two problems exactly using trust-region methods discussed in Algorithm section.

In order to guarantee convergence of the univariate function to the original optima, it is critical to determine when the function is twice continuously differentiable. It is also important to find a strategy that provides a good initial guess to solve P4. The answer to these vital questions are provided in the next section.

Theoretical Developments

We realize that the lower level sub-problem (P4) can be interpreted as a parametric problem in t and therefore, we can utilize the optimal value function properties

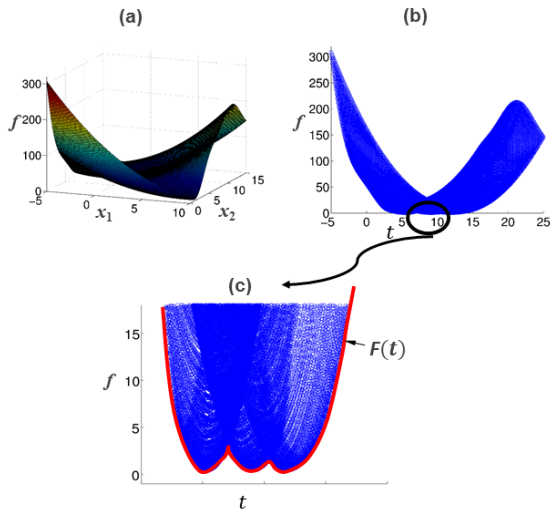

Figure 1. (a) Branin function (Global minima: $f(x^{opt}) = 0.397887$ at $x^{opt} = (-\pi, 12.275)$, $(\pi, 12.275)$, $(9.42478, 2.475)$), (b) Projection on t-space, (c) Illustrating the minima of projection.

discussed in Fiacco (1984) for general nonlinear parametric problems.

Properties of the projected function F(t)

We list important properties of $F(t)$:

Property 1. *$F(t)$ is continuous if $f(x)$ is continuous.*

Property 2. *$F(t)$ is convex (concave) if $f(x)$ is convex (concave).*

Property 3. *Assume that KKT conditions, Second Order Sufficient Condition (SOSC), Linear Independence Constraint Qualification (LICQ) and Strict Complementarity Slackness (SCS) hold true at $\bar{x}_p$ (optima corresponding to $t = t_p$) with the associated Lagrange multipliers $\bar{\nu}_1$, $\bar{\nu}_2$ and $\bar{\lambda}$, then in the neighborhood of t_p, $F(t)$ is twice continuously differnentiable.*

Property 4. $F(t^L) = f(x^L), \quad F(t^U) = f(x^U)$

While Property 4 is trivial to prove, Property 1, 2, 3 can be shown through the works of Hogan (1973), Fiacco and Kyparisis (1986) and Fiacco (1984) respectively. Note that Property 3 will not hold whenever the assumptions are violated and convergence may not be guaranteed.

Obtaining the initial guess to solve P4

Theorem 1. *(Fiacco (1984))*
Suppose assumptions of Property 3 hold true, then (i) $\bar{x}_p$ is a local minimizer and the Lagrange multipliers are unique (ii) there exists a once continuously differentiable vector function $y = [x(t), \nu_1(t), \nu_2(t), \lambda(t)]$ such that $x(t)$ satisfies the KKT conditions of P4 at t in the neighborhood of t_p:

$$\frac{dy(t)}{dt} = M^{-1}N \qquad \textit{where} \tag{6}$$

$$M = \begin{bmatrix} \nabla^2 L & \nabla g_1^{(1)} \cdots & \nabla g_n^{(1)} & \nabla g_1^{(2)} & \cdots & \nabla g_n^{(2)} & \nabla h_1 \\ -\nu_1^{(1)} \nabla^T g_1^{(1)} & -g_1^{(1)} & & & & & \\ \vdots & \ddots & & & & & \\ -\nu_n^{(1)} \nabla^T g_n^{(1)} & & -g_n^{(1)} & & & & \\ -\nu_1^{(2)} \nabla^T g_1^{(2)} & & & -g_1^{(2)} & & & \\ \vdots & & & & \ddots & & \\ -\nu_n^{(2)} \nabla^T g_n^{(2)} & & & & & -g_n^{(2)} & 0 \\ \nabla^T h_1 & \cdots & & 0 & & & \end{bmatrix}$$

$$N = \left[\nabla_{tx}^2 L, -\nu_1^{(1)} \nabla_t^T g_1^{(1)}, \ldots - \nu_n^{(2)} \nabla_t^T g_n^{(2)}, \nabla_t^T h_1\right]^T$$

A first order Taylor series approximation can be used to obtain the optimal solution at a point t_{p+1} in the neighborhood of t_p by (Fiacco (1984)):

$$\begin{bmatrix} x(t_{p+1}) \\ \nu^{(1)}(t_{p+1}) \\ \nu^{(2)}(t_{p+1}) \\ \lambda(t_{p+1}) \end{bmatrix} = \begin{bmatrix} \bar{x}_p \\ \bar{\nu_p}^{(1)} \\ \bar{\nu_p}^{(2)} \\ \bar{\lambda_p} \end{bmatrix} + \bar{M}^{-1} \bar{N} (t_{p+1} - t_p) \tag{7}$$

Theorem 1 has been used to obtain analytical solutions for mp-QPs (Dua et al., 2002). In the case of classical nonlinear programming, where the Hessian and the Jacobian are analytically available, Theorem 1 in conjunction with first-order approximation can yield an exact optimum in the neighborhood of t_p. Property 4 and Eq. 7 give a powerful tool to approximate $F(t)$ starting from the lower bound (t^L) without performing optimization. In other words, there exists a projection/map given by Theorem 1 which can convert the n-dimensional optimization problem to a 1-dimensional problem.

Note that Eq. 7 requires the Hessian and the Lagrange multipliers corresponding to $\bar{x}_p$. In the paradigm of derivative-free optimization, obtaining an analytical expression for the Hessian is not possible and using finite difference is computationally expensive. A class of fully-quadratic models (Conn et al., 2009), approximating the Hessian of the original function f, can be constructed but that requires evaluations to the order of n^2. Additionally, the result is applicable for parameters in the neighborhood of t_p and taking small steps in t may

not be practical. Nevertheless, Eq. 7 provides a good initial guess for a trust-region method that converges to the optimum, $\bar{x}_p$.

DFO Algorithm

The idea is to partition t into M_t segments and obtain a finite number of samples along t $(t_1, \ldots, t_{M_t+1})$ and the corresponding $F(t)$ $(F(t_1), \ldots, F(t_{M_t+1}))$ and transform back the optima of the univariate function, t^* to the original space, x^*.

The algorithm has two components:
a) An inner loop - finds optima $(F(t_p), \bar{x}_p)$ of P4 at the sampled parameter value t_p with starting point $x_{0,p}$ using an n-dimensional surrogate model ,
b) An outer loop - utilizes function values F obtained from the inner loop to train a 1-dimensional surrogate model and obtain t^*.

Solving P4 (Inner Loop)

The steps of algorithm described in Bajaj and Hasan (2016) are followed to obtain $\bar{x}_p$, given an initial guess $x_{0,p}$. Instead of using the derivative of the penalty function as the criticality measure, we solve Eq. 8 and stop when both the KKT error (defined by Eq. 8) , $\epsilon_{t,p} = \sum_i^n SP_i + SN_i$ and the accuracy of the surrogate model $\epsilon_{f,p}$ are less than their respective tolerances, ϵ_t and ϵ_f.

The following linear program is solved to check whether a point $x_{c,p}$ is an optima of P4 and determine the Lagrange multipliers that satisfy the KKT conditions:

$$\begin{aligned} \min_{\substack{SP_i, SN_i, \\ \nu^{(1)}, \nu^{(2)}, \lambda}} \quad & \sum_{i=1}^{n} SP_i + SN_i \qquad (8) \\ s.t. \quad & \frac{\partial f^r}{\partial x_i} + \lambda - \nu_i^{(1)} + \nu_i^{(2)} = SP_i - SN_i \\ & \nu^{(1)}(x^L - x_{c,p}) = 0, \quad \nu^{(2)}(x_{c,p} - x^U) = 0, \\ & \nu^{(1)}, \nu^{(2)} \geq 0, \quad x^L \leq x_{c,p} \leq x^U \end{aligned}$$

where $f^r : \mathbb{R}^n \to \mathbb{R}$ represents a surrogate model of f. The satisfaction of above equation is not sufficient to guarantee optimality since the surrogate model may be inaccurate. Therefore, $x_{c,p}$ has to satisfy the KKT conditions and the surrogate model needs to be accurate in the current trust region $\Delta_{k,p}$ in order to guarantee convergence to the actual optima $\bar{x}_p$. When both these conditions are satisfied, $x_{c,p}$ is the optima of P4 i.e. $x_{c,p} = \bar{x}_p$. The Lagrange multipliers are recorded and the Hessian of the surrogate model is used as an estimate of the Hessian of the original function. Since this procedure provides only an estimate of Hessian and multipliers and a large step size is used, applying Eq. 7 may not give the exact optima,$\bar{x}_{p+1}$. But still, it gives a good initial guess $(x_{0,p+1})$ for an algorithm that converges to local optima $(\bar{x}_{p+1})$.
Cubic radial basis function is used as the surrogate model since it has been shown to satisfy the fully-linear property (Wild and Shoemaker (2013)). We use global optimization solver ANTIGONE (Misener and Floudas, 2014) while optimizing f^r.

Solving P5 (Outer loop)

(i) Initialize the algorithm by splitting the entire t domain into M_t segments such that $t_1 = t^l$ and $t_{M_t+1} = t^u$.
(ii) Use the inner loop algorithm for each $t \in \{t_1, \cdots, t_p, \cdots t_{M_t+1}\}$ to obtain a set of values $\{F(t_1), \cdots F(t_p), \cdots F(t_{M_t+1})\}$ and record the optima $\{\bar{x_1}, \cdots, \bar{x_p}, \cdots \bar{x}_{M_t+1}\}$.
(iii) Construct a single dimensional surrogate model using the samples obtained above (Assumption of fully-linear property will hold due to Theorem 1 and Theorem 3).
(iv) Estimate the root mean squared error $\epsilon_{l,o}$.
(v) Globally optimize $F^r(t)$ to obtain t_d and use the inner loop algorithm to obtain the optimum $\bar{x_d}$.
(vi) If $\epsilon_{l,o} \leq \epsilon_f$ and $\nabla F^r(t_d) \leq \epsilon_t$, $t^* = t^d$ and $x^* = \bar{x}_d$, STOP. Else, go to step (vii).
(vii) If $F(t_d) \leq F(t_l)$, set $x_{l+1} = x_d$ and increase the trust region $\Delta_{l+1} = \sigma_2 \Delta_l$.
(viii) Else set $x_{l+1} = x_l$ and decrease the trust region $\Delta_{l+1} = \sigma_1 \Delta_l$.
(ix) Use the previously obtained L_p points in Δ_{l+1} and divide the new domain into $M_t - L_p$ segments and go to step (ii).

Computational Results

The proposed optimization algorithm is applied on a test suite of 17 problems from literature (Rios and Sahinidis (2013) and Wild and Shoemaker (2013)). All the test problems are box-constrained nonconvex problems. It is assumed that only function evaluation at a desired sample point is possible. The parameter values: $\eta_0, \sigma_1, \sigma_2, \epsilon_t, \epsilon_f, M_t$ are chosen to be 0.25, 0.5, 3, 0.1, 0.01 and 20 respectively. The maximum number of function evaluations, N_{max} are limited to 2500. The

Table 1. Summary of results

Problem	N_{var}	f^{opt}	N^{p}_{eval}	N^{ASA}_{eval}	N^{DD}_{eval}	N^{DE}_{eval}	N^{NEW}_{eval}
Branin	2	0.3978	281	2517[1]	133	1128	32
Hartman 3	3	-3.862	576	-	-	-	-
Shekel 5	4	-10.153	665	2538[1]	253[1]	2500[1]	15
Shekel 7	4	-10.402	1004	-	-	-	-
Shekel 10	4	-10.5364	803	-	-	-	-
Ex8.1.1	2	-2.021	234	2515	91	1685	91
Schwefel	5	0	501	2551	2423	59	183
Camel 6	2	-1.031	441	2517	193	2098	40
Camel 1	2	-1.031	527	2517	129[1]	2251	35
Hs 5	2	-1.913	307	2517	121	554	33
Hs 045	5	1	1350	2551[1]	246	289[1]	164[1]
Himmelp1	2	-62.053	244	2514	142	1428	34
Powell	4	0	2500[2]	2538[1]	354[1]	543[1]	379
Kowalik	4	3.07E-4	822	2521	1583	786	195
Hatfldc	4	0	2347	2538[1]	794[1]	828[1]	1372
Hartman 6	6	-3.322	2500[2]	2551[1]	1559	1214[1]	127
Trigono	5	0	2500[2]	-	-	-	-

results of the test problems are given in Table 1. Column 1 lists test problems; column 2 gives the number of variables, N_{var}; column 3 lists the global optima of each problem, f^{opt}; column 4 provides the number of evaluations taken by our proposed method to converge to the optimal solution. The columns 5, 6, 7, 8 and 9 record the number of evaluations taken by other methods, namely the Adaptive Simulated Annealing (ASA), DAKOTA/DIRECT (DD), DAKOTA/EA (DE), TOMLAB/GLCCLUSTER (TG) and NEWUOA (NEW), respectively. The results of the other methods are taken from http://archimedes.cheme.cmu.edu/?q=dfocomp, where Rios and Sahinidis (2013) compared 22 solvers by giving 10 runs with different initial guesses for each test problem. The numbers reported in Table 1 is the minimum of the number of evaluations taken to achieve the best solution in 10 runs.

In all the problems, the proposed algorithm either converges to the global optima or is very close to it given the budget of function evaluations. The proposed method performs better in terms of number of evaluations and number of problems converging to global optima compared to ASA. In 7 out of 13 problems, DD performs better; while in 4 of the problems, our method gives a better objective function value and in 2 of the problems, we perform better in terms of number of evaluations relative to DD. Compared to DE, overall, our method achieve superior results, while NEWUOA seems to perform better. However, performance of all the solvers except DD compared in this paper depends heavily on the initial guess provided. For instance, in one of the runs of NEWUOA on "hatfldc", the solver converges to $f(x^*) = 1.56 \times 10^9$, while the optimum function value is 0. This work proposed a method which is invariant to initial guess. We also cover the entire domain in a specific direction and increase chances of identifying the global optima.

We have noticed that the majority of the evaluations are spent in the inner loop (refining $F(t)$) because the initial guess provided might not be "close enough". We note that, even though it is theoretically possible to obtain the optimum, $\bar{x}_{p+1}$ corresponding to t_{p+1} using $\bar{x}_p$, it requires the Hessian and the multiplier information, and the step size to be small. It remains to be examined if the number of evaluations needed to provide better Hessian and multiplier estimates would offset the function calls required in the inner loop.

[1]Optima was not obtained

[2]Maximum evaluations allowed

Concluding remarks

We proposed a novel method to optimize an n-dimensional problem by projecting on a special 1-dimensional space and showed that the optima of a special 1-dimensional function corresponds to the optima of the original function. We then showed that, using results from sensitivity analysis literature, the 1-dimensional function can be approximated. A two-step trust-region based algorithm was proposed where we iteratively used n-dimensional and 1-dimensional surrogate models to converge to a KKT point. The efficacy of the method was illustrated by successfully applying on 17 nonconvex test problems. Further investigation is needed for the inner loop algorithm to reduce the overall evaluations.

This work provides a unique perspective to solve black-box problems. The framework is generic and can be extended to constrained black-box problems. In the constrained case, Theorem 1 can still be used to provide initial guess and inner loop can be replaced by an alternate algorithm such as filter technique (Eason and Biegler (2016)) to handle black-box constraints.

Acknowledgments

Financial support from Texas A&M Energy Institute is gratefully acknowledged.

References

Bajaj, I. and Hasan, M. M. F. (2016). Effective sampling, modeling and optimization of constrained black-box problems. *Computer Aided Process Engineering*, 38:553–558.

Boukouvala, F., Hasan, M. M. F., and Floudas, C. A. (2015). Global optimization of general constrained grey-box models: new method and its application to constrained PDEs for pressure swing adsorption. *Journal of Global Optimization*, pages 1–40.

Caballero, J. A. and Grossmann, I. E. (2008). An algorithm for the use of surrogate models in modular flowsheet optimization. *AIChE Journal*, 54(10):2633–2650.

Conn, A. R., Scheinberg, K., and Vicente, L. N. (2009). Global convergence of general derivative-free trust-region algorithms to first-and second-order critical points. *SIAM Journal on Optimization*, 20(1):387–415.

Dua, V., Bozinis, N. A., and Pistikopoulos, E. N. (2002). A multiparametric programming approach for mixed-integer quadratic engineering problems. *Computers & Chemical Engineering*, 26(4):715–733.

Eason, J. P. and Biegler, L. T. (2016). A trust region filter method for glass box/black box optimization. *AIChE Journal*.

Fiacco, A. V. (1984). *Introduction to sensitivity and stability analysis in nonlinear programming*. Academic Press, New York.

Fiacco, A. V. and Kyparisis, J. (1986). Convexity and concavity properties of the optimal value function in parametric nonlinear programming. *Journal of optimization theory and applications*, 48(1):95–126.

First, E. L., Hasan, M. M. F., and Floudas, C. A. (2014). Discovery of novel zeolites for natural gas purification through combined material screening and process optimization. *AIChE Journal*, 60(5):1767–1785.

Hasan, M. M. F., First, E. L., and Floudas, C. A. (2013). Cost-effective CO_2 capture based on in silico screening of zeolites and process optimization. *Physical Chemistry Chemical Physics*, 15(40):17601–17618.

Henao, C. A. and Maravelias, C. T. (2011). Surrogate-based superstructure optimization framework. *AIChE Journal*, 57(5):1216–1232.

Hogan, W. W. (1973). Point-to-set maps in mathematical programming. *Siam Review*, 15(3):591–603.

Misener, R. and Floudas, C. A. (2014). ANTIGONE: Algorithms for coNTinuous/Integer Global Optimization of Nonlinear Equations. *Journal of Global Optimization*, 59(2-3):503–526.

Nuchitprasittichai, A. and Cremaschi, S. (2011). Optimization of CO_2 capture process with aqueous amines using response surface methodology. *Computers & Chemical Engineering*, 35(8):1521–1531.

Powell, M. J. (2006). The NEWUOA software for unconstrained optimization without derivatives. In *Large-scale nonlinear optimization*, pages 255–297. Springer.

Rios, L. M. and Sahinidis, N. V. (2013). Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293.

Shan, S. and Wang, G. G. (2010). Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. *Structural and Multidisciplinary Optimization*, 41(2):219–241.

Wild, S. M. and Shoemaker, C. (2013). Global convergence of radial basis function trust-region algorithms for derivative-free optimization. *SIAM Review*, 55(2):349–371.