
Thermoacoustic stabilization of combustors with gradient-augmented Bayesian optimization and adjoint models

Symposium on Thermoacoustics in
Combustion: Industry meets Academia
(SoTiC 2021)
Sept. 6 - Sept. 10, 2021
Munich, Germany
Paper No.: 8476
©The Author(s) 2021

Ushnish Sengupta¹ and Matthew P. Juniper¹

Abstract

Bayesian optimization (BO) is a global optimization algorithm well-suited for multimodal functions that are costly to evaluate, e.g. quantities derived from computationally expensive simulations. Recent advances have made it possible to scale BO to high-dimensional functions and accelerate its convergence by incorporating derivative information. These developments have laid the groundwork for a productive interplay between BO and adjoint solvers, a tool to cheaply obtain gradients of objective functions w.r.t. tunable parameters in a simulated physical system. In thermoacoustics, adjoint-based optimization has previously been applied to Helmholtz solvers and low-order network models to find optimally stable combustor configurations. These studies have used conjugate gradient or quasi-Newton optimizers which can get stuck in local optima and may require many evaluations of the underlying model to find a good optimum. In this paper, we propose using gradient-augmented BO to optimize adjoint models. We consider two test cases from the thermoacoustics literature: optimizing design parameters in a 1D adjoint Helmholtz model of a Rijke tube and geometry optimization in a low-order network model of a longitudinal combustor. We show that compared to BFGS, a standard quasi-Newton method, our gradient-enhanced BO arrives at multiple, more optimal configurations using considerably fewer evaluations of the solver. This approach holds great promise for efficient thermoacoustic stabilization when designing using expensive 3D adjoint Helmholtz solvers.

Keywords

thermoacoustic instabilities, combustion, adjoint models, Bayesian optimization

Introduction

Thermoacoustic oscillations pose a major challenge to designers of gas turbines, jet engines and rockets. These instabilities arise from positive feedback loops between fluctuations in the heat release rate and acoustic waves in the combustor. They may result in unacceptable levels of noise, enhanced heat transfer through the chamber walls and structural damage.

Thermoacoustic stability is exceedingly sensitive to changes in the acoustic characteristics of a combustion chamber and the acoustic/ hydrodynamic response of the flame. As a result, small changes in the system geometry, boundary conditions or flame behavior can often stabilise an unstable combustor. However, coming up with a suitable design modification through pure trial-and-error experimentation is usually infeasible. A famous example is the F1 engine of the Saturn V rocket whose stabilization required 2000 full-scale engine tests; baffles were introduced in the injector plate to suppress thermoacoustic oscillations [Joos \(2006\)](#). Adjoint-based sensitivity analysis in thermoacoustic models offers a systematic procedure for discovering appropriate design changes.

The advantage of adjoint methods lies in the efficient computation of derivatives of a given objective function at a cost that scales independently of the number of design variables. The sensitivity of eigenvalues (natural frequencies and growth rates) of a thermoacoustic model w.r.t. to design variables can therefore be computed cheaply using adjoint methods; this information is then utilized by an optimization routine from the gradient descent family to arrive at a stabilized design. Adjoint-based sensitivity analysis has been applied to low-order network models by [Magri and Juniper \(2013\)](#), [Mensah and Moeck \(2017\)](#), [Silva et al. \(2017\)](#), to 2D Helmholtz equation models by [Falco and Juniper \(2021\)](#) and to 3D Helmholtz models by [Mensah et al. \(2019\)](#). A comprehensive review of the use of adjoints in thermoacoustic can be found in [Magri \(2019\)](#).

Other techniques have also been used to stabilize designs. [Aguilar Perez \(2019\)](#) used an exhaustive grid search while [Jones et al. \(2021\)](#) used a genetic algorithm to optimize the shape of a thermoacoustically unstable combustor. Compared to gradient descent using adjoints, these tools have the advantage of not getting trapped in local minima, but are much slower in comparison and require many expensive simulations to be run for each configuration they evaluate.

¹Department of Engineering, University of Cambridge, UK

Corresponding author:

Matthew Juniper, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK.
Email: mpj1001@cam.ac.uk

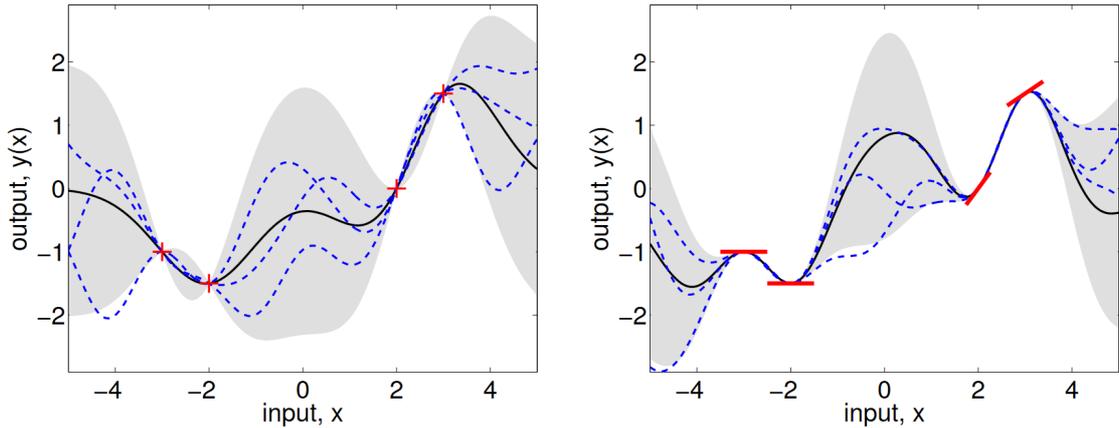


Figure 1. A 1D Gaussian Process metamodel with and without derivatives. Including derivative information results in a GP which has a higher confidence in its predictions. The gray area represents 3 s.d. uncertainty bounds in the Gaussian process metamodel, the red crosses are function evaluations, the red lines denote function and derivative evaluations and the dotted blue lines are randomly sampled functions from the Gaussian process metamodel.

In this paper, we propose using gradient-augmented Bayesian optimization (BO) to optimize adjoint models. BO is able to find global optima in expensive, multi-modal functions with strikingly few function evaluations. It does so by building a Gaussian Process (GP) metamodel of the objective function at every iteration. To decide where to evaluate the objective function next, an acquisition function such as expected improvement [Huang et al. \(2006\)](#), upper confidence bound [Srinivas et al. \(2009\)](#) or knowledge gradient [Scott et al. \(2011\)](#) is used. This balances exploration (i.e., moving to places where the GP has high predictive uncertainty) and exploitation (i.e., moving to where the GP has a high predictive mean). Vanilla BO, however, does not utilize derivative information. Nevertheless, since the derivative of a GP with a twice-differentiable kernel is also a GP [Williams and Rasmussen \(1996\)](#), it is possible to model derivatives in a GP metamodel. The incorporation of derivative observations in the GP model leads to lower predictive variances (Figure 1 shows a simple example of a 1D GP with and without derivatives) and makes our exploration of the search space much more efficient. The main challenge is the impractical $\mathcal{O}(n^3 D^3)$ computational cost of building a GP model with n function evaluations and D derivatives. In this paper, we use a scalable gradient-augmented Bayesian optimization algorithm loosely based on [Eriksson et al. \(2018\)](#).

To demonstrate the effectiveness of gradient-augmented Bayesian optimization, we consider two optimization problems from the literature on adjoint models in thermoacoustics. The first one is a simple toy problem from [Juniper \(2018\)](#) where the iris diameter and heater position in a 1D Helmholtz

equation model of a Rijke tube must be optimized. The second involves geometry optimization in a low-order network model of a longitudinal combustor from [Aguilar and Juniper \(2020\)](#). We compare gradient-augmented BO with BFGS, a standard quasi-Newton optimizer which uses an estimate of the inverse Hessian matrix to improve convergence. Compared to BFGS, we find that the gradient-augmented BO does not get stuck in local optima and requires fewer evaluations of the solver to arrive at more thermoacoustically stable configurations. It is also able to efficiently explore the whole design space and locate multiple suitable combinations of design parameters, which is useful further downstream in the design cycle.

Methods

Thermoacoustic models

Our goal is to design systems in which all eigenmodes decay in time, so we use the summed exponential of the growth rates as our objective function. This strongly penalizes positive growth rates, but the rewards for stabilizing the system diminish progressively as growth rates become less than zero. If \mathbf{X} is the vector that contains all the combustor parameters and λ_i -s are growth rates of the eigenmodes of interest, then our cost function is given by

$$\mathcal{J}(\mathbf{X}) = \sum_i e^{\lambda_i} \quad (1)$$

We only consider longitudinal thermoacoustic modes in this paper, however we note that the above expression would result in double-counting of symmetric azimuthal modes in a system that has them. When optimizing such a system, the corresponding terms should be weighed by 0.5 to avoid penalizing them doubly.

The first test case considered in this paper comes from [Juniper \(2018\)](#) where optimization using adjoints is demonstrated on a 1D Helmholtz model of a Rijke tube. A finite element discretization of the weak form is used and the discrete adjoint framework is used to compute eigenvalue sensitivities. The design parameters are the heater position and the diameter of a variable-diameter iris placed at the downstream boundary. The growth rate of the fundamental mode, which we seek to minimize, has two local minima (Figure 2) in the search space.

The second test case is taken from [Aguilar and Juniper \(2020\)](#) and relates to geometry optimization in a network model of Rama Balachandran's 10 kW longitudinal combustor built in Cambridge, originally intended for the experimental investigation of the response of turbulent premixed flames to acoustic oscillations [Balachandran et al. \(2005\)](#). The geometry consists of an inlet duct connected to a plenum with a linearly varying cross section on either end. This leads to the neck which contains the fuel injection plane and a centred bluff body used to stabilize the flame. The outlet is a cylindrical pipe which contains

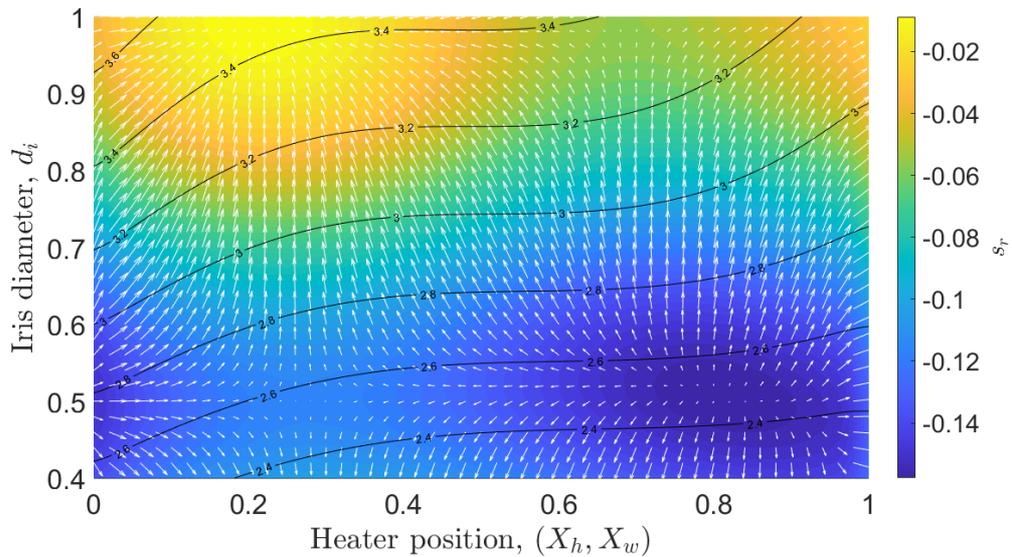


Figure 2. Contour plots of non-dimensionalized growth rate and frequency of the Rijke tube fundamental mode plotted against heater position and iris diameter, reproduced from Juniper (2018). The numbered black contour lines correspond to frequencies and the colored contour plot represents the growth rate. White arrows indicate the direction and magnitude of the growth rate gradients, as computed by the adjoint model.

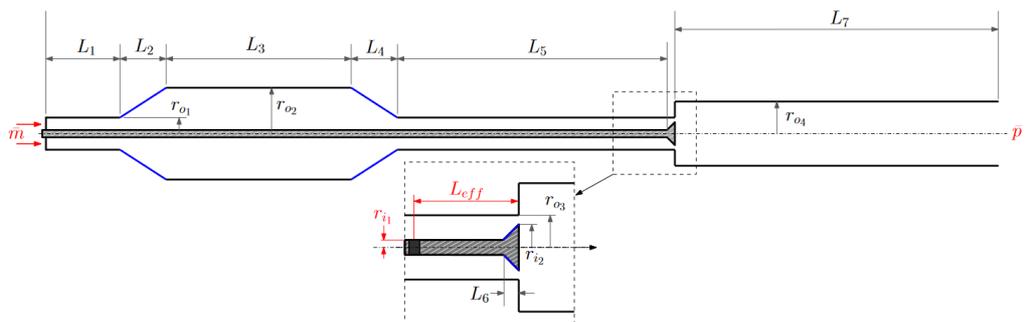


Figure 3. Geometric parameters of the network model from Aguilar Perez (2019) used in the optimization routine. The parameters in black are allowed to be updated. The parameters in red are kept constant.

the flame. Figure 3 shows a schematic of the combustor model and the geometric parameters being optimized. The rig is modeled using a one-dimensional network model with 124 straight ducts. This model has 4 unstable eigenmodes between 0 and 1000 Hz that need to be stabilized with geometry modifications. We restrict the search space to a box bounded by $\pm 4.5\%$ of the areas and lengths in the original configuration.

For more details on the test cases, we refer the interested reader to the original papers.

Gradient-augmented Bayesian optimization

Bayesian optimization is a powerful tool but the inclusion of D extra pieces of gradient information per point causes severe scaling issues, particularly when a large number of iterations are run. In this paper, we use the algorithm from [Eriksson et al. \(2018\)](#) for scaling GPs with derivatives. The cubic scaling with the number of functional evaluations for kernel learning in GPs stems from the fact that a linear system must be solved, which necessitates a Cholesky factorization of the kernel matrix. To work around this, we approximate the true kernel with a structured kernel interpolation for products (D-SKIP) approximation from [Gardner et al. \(2018\)](#) which allows fast matrix-vector multiplication. The structured kernel interpolation (SKI) approach uses local polynomial interpolation on an induced grid with sparse weights to approximate the kernel matrix in each dimension. Instead of computing kernel values between data points directly, SKI computes kernel values between inducing points and interpolates these kernel values to approximate the true data kernel values $k(x, x') \approx \sum_i w_i(x)k(x_i, x')$. The final kernel matrix can be written as a Hadamard product (\odot) of approximate one-dimensional matrices; for example, in two dimensions, the kernel matrix with derivatives is approximated as follows:

$$K^\nabla \approx \begin{bmatrix} W_1 K_1 W_1^T & W_1 K_1 \partial W_1^T & W_1 K_1 W_1^T \\ \partial W_1 K_1 W_1^T & \partial W_1 K_1 \partial W_1^T & \partial W_1 K_1 W_1^T \\ W_1 K_1 W_1^T & W_1 K_1 \partial W_1^T & W_1 K_1 W_1^T \end{bmatrix} \odot \begin{bmatrix} W_2 K_2 W_2^T & W_2 K_2 W_2^T & W_2 K_2 \partial W_2^T \\ W_2 K_2 W_2^T & W_2 K_2 W_2^T & W_2 K_2 \partial W_2^T \\ \partial W_2 K_2 W_2^T & \partial W_2 K_2 W_2^T & \partial W_2 K_2 \partial W_2^T \end{bmatrix} \quad (2)$$

where W_j and K_j denote the structured kernel interpolation (SKI) and inducing point grid matrices in the j -th coordinate direction. Additionally, we use the iterative conjugate gradient method with preconditioning for solving the linear system instead of Cholesky factorization. These tricks enable $\mathcal{O}(nD)$ complexity kernel learning for the GP, where n is the number of data points and $\mathcal{O}(1)$ prediction per test points.

Once we have the GP surrogate model of our function, we use its mean and uncertainty to compute the expected improvement acquisition function ([Jones et al. \(1998\)](#)) which helps us decide where best to sample the function and its derivative in each iteration. Expected improvement is defined as

$EI(\mathbf{x}) = \mathbb{E}[\max(f(\mathbf{x}^+) - f(\mathbf{x}), 0)]$ where $f(\mathbf{x}^+)$ is the value of the best sample so far and \mathbf{x}^+ is the location of that sample i.e. $\mathbf{x}^+ = \operatorname{argmax}_{\mathbf{x}_i \in \mathbf{x}_{1:t}} f(\mathbf{x}_i)$. The expected improvement can be evaluated analytically for a Gaussian process:

$$EI(\mathbf{x}) = \begin{cases} (f(\mathbf{x}^+) - f(\mathbf{x}) - \xi)\Phi(Z) + \sigma(\mathbf{x})\phi(Z) & \text{if } \sigma(\mathbf{x}) > 0 \\ 0 & \text{if } \sigma(\mathbf{x}) = 0 \end{cases} \quad (3)$$

where

$$Z = \begin{cases} \frac{f(\mathbf{x}^+) - \mu(\mathbf{x}) - \xi}{\sigma(\mathbf{x})} & \text{if } \sigma(\mathbf{x}) > 0 \\ 0 & \text{if } \sigma(\mathbf{x}) = 0 \end{cases} \quad (4)$$

$\mu(\mathbf{x})$ and $\sigma(\mathbf{x})$ are the mean and the standard deviation of the GP posterior predictive at \mathbf{x} , respectively. Φ and ϕ are the CDF and PDF of the standard normal distribution, respectively. The first summation term in Equation (3) is the exploitation term which encourages the algorithm to explore where the surrogate model has a lower predictive mean than the current best sample and the second term is the exploration term which encourages sampling new points in regions of high uncertainty.

The parameter ξ in Equation (3) determines the amount of exploration during optimization and higher ξ values lead to more exploration. With increasing ξ values, the importance of improvements predicted by the GP posterior mean $\mu(\mathbf{x})$ decreases relative to the importance of potential improvements in regions of high prediction uncertainty, represented by large $\sigma(\mathbf{x})$ values. We use the default value $\xi = 0.01$ recommended by [Lizotte \(2008\)](#).

In most high-dimensional optimization problems of practical significance, there exist only a few relevant directions that capture most of the variation in the function. This is exploited by the active subspace dimension reduction method [Constantine et al. \(2014\)](#). The optimal subspace is given by the dominant eigenvectors of the covariance matrix $C = \int_{\Omega} \nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T$, estimated by Monte Carlo integration. Because we already have access to gradient information, it is natural to combine active subspaces with BO to reduce the computational overhead of building our GP metamodel at every iteration. Random projections (without gradients) have already been used in the BO literature for high-dimensional functions [Wang et al. \(2013\)](#). In this paper, we learn the active subspace that captures 95% of the variation at every iteration, choose d random directions of the active subspace onto which the function is to be projected and add a small amount of random noise to the projection matrix. The subspace

projection is not used for the 1D Rijke tube problem, since the optimization problem is two-dimensional. $d = 3$ is used for the longitudinal combustor optimization.

Algorithm 1: BO with derivatives and active subspace learning

sample cost function f and gradients ∇f at D initial points using simple Latin Hypercube sampling, where D is the dimension of the design space.

while *computational budget remaining* **do**

Estimate active subspace of cost function using sampled gradients that capture 99% of the variation.

Pick d directions at random from estimated active subspace to compute projection $P \in \mathbb{R}^{D \times d}$.

Add a small amount of random Gaussian noise to obtain the final projection matrix P_{noisy} .

Update hyperparameters of GP with gradient defined by kernel $k(P_{noisy}^T \mathbf{x}, P_{noisy}^T \mathbf{x}')$ and construct the GP surrogate.

Use the GP surrogate to optimize the expected improvement $\mathbf{u}_{n+1} = \operatorname{argmax} \operatorname{EI}(\mathbf{u})$ to obtain the next candidate point $\mathbf{x}_{n+1} = P_{noisy} \mathbf{u}_{n+1}$.

Sample point \mathbf{x}_{n+1} , value f_{n+1} , and gradient ∇f_{n+1}

Update data $D_{n+1} = D_n \cup \{x_{n+1}, f_{n+1}, \nabla f_{n+1}\}$

end

Results

For the 1D Rijke tube problem, the gradient-augmented BO consistently finds the known global optimum near $(0.82, 0.5)$ fairly rapidly across all trials, whereas a gradient-descent-based approach like BFGS predictably converges to either of the two local optima (the other one is located at $(0.0, 0.5)$), depending on the starting point chosen. Figure 4 illustrates how the gradient-augmented Bayesian optimization refines its surrogate model with increasing number of iterations. After 7 evaluations, the surrogate is very inaccurate because the bottom right and top left quadrants are left unexplored. However, these regions also have higher uncertainty, so we observe that by 17 evaluations, these regions have also been explored. After this exploration phase, the BO deems that the model is accurate enough for the purpose of optimization and spends the next 10 evaluations trying to converge on the global optimum. BO consistently explores both local minima regardless of initialization. This is very useful in thermoacoustics because almost all design optimization problems have multiple stable configurations. Discovering more of them makes the designer's life easier, since thermoacoustic stability is typically but one of many conflicting design objectives. The MATLAB code for the 1D Rijke tube optimization may be found [here](#).

To compare between gradient-augmented BO and BFGS, we conduct 10 trials of both algorithms on the longitudinal combustor test problem. In each trial, the initial points are sampled uniformly from

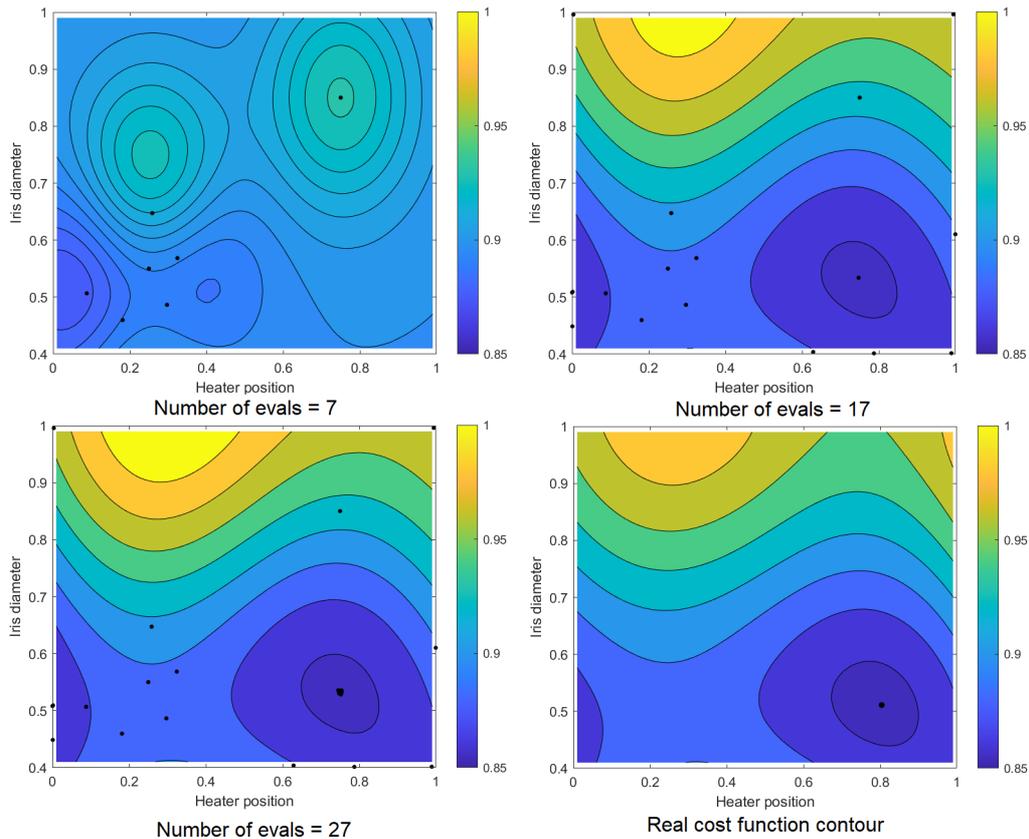


Figure 4. Surrogate Gaussian process model of the Rijke tube cost function at different iterations. Black dots indicate where the Bayesian optimization has evaluated the function and its gradients. The bottom right plot is a contour plot of the true cost function, with the black dot indicating the location of the true global minimum.

the design space with a different seed for the pseudo-random number generator. When the number of iterations is very small, the BFGS appears to outperform Bayesian optimization, but this is because Bayesian optimization explores more early on due to the high initial uncertainty in the Gaussian Process surrogate. BO quickly overtakes the BFGS average, however, and by the 50th iteration, 9 out of 10 BO runs have found a configuration as stable as the best BFGS run. The final average cost function after 50 runs is 1.62 for BFGS and 1.01 for BO. As in the Rijke tube case, multiple distinct stable geometries are found in each run. In general, stable configurations have a reduced flame holder area and a reduced neck area, because the eigenvalues are quite sensitive to these two parameters. Figure 6 shows a stable geometry found by one of the BO runs.

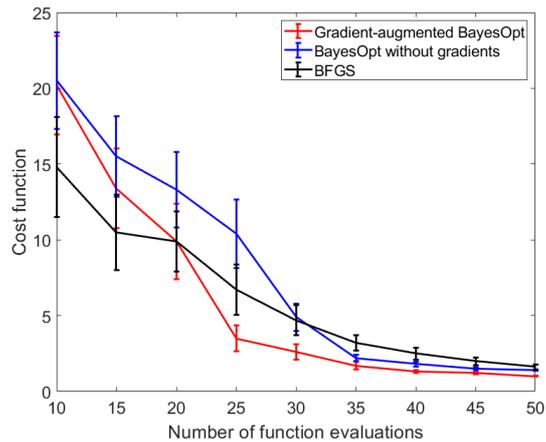


Figure 5. Best objective function values achieved by the BFGS, BO without gradient and gradient-augmented BO routines in the longitudinal combustor test case, averaged across 10 trials, plotted against the number of cost function evaluations. Errorbars indicate 1 s.d. variation across trials.

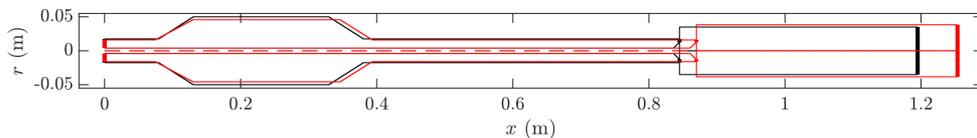


Figure 6. A stable combustor configuration, shown in red, found by a gradient-augmented Bayesian optimization run after 50 evaluations. Original unstable configuration in black.

Figure 5 also shows the comparison between the gradient-augmented BO to BO without gradient values. As expected, it converges to the global optimum more slowly. This shows the benefit of having derivative information from the adjoint models.

To be fair to BFGS, it should be pointed out the BO involves a modest computational overhead whereas BFGS has nearly none. For the longitudinal combustor model, it takes around 30 seconds per iteration for the gradient-augment BO on a Dell G7 7590 laptop with a 4 core 8th Generation Intel i5-8300HQ processor. However, when optimizing expensive models with this algorithm, the evaluation of the cost function and its gradient becomes the computational bottleneck while the cost per BO iteration becomes a non-issue.

Conclusions

In this study, we demonstrate that scalable, gradient-augmented BO pairs perfectly with adjoint methods in thermoacoustics. We use two optimization problems from the literature as our test cases: a 1D Helmholtz equation model of a Rijke tube and a low-order network model of a longitudinal combustor. We showed that the BO algorithm consistently finds more stable parameter combinations using fewer iterations than BFGS, a popular quasi-Newton optimizer. It also builds a metamodel of the objective function in the entire design space and explores multiple feasible designs. The metamodel can naturally be used for uncertainty quantification as well.

There are several potential avenues for future work. While this paper illustrates the untapped potential of gradient-augmented BO in thermoacoustics and other engineering disciplines where adjoint models are commonly used, the models we used in our test cases were relatively cheap to evaluate. BO really shines when the objective is expensive to evaluate, so a truly convincing use case would involve more expensive models, e.g. 3D adjoint Helmholtz solvers and a complex combustor geometry with many hundreds of parameters. Another interesting line of research would be looking at different acquisition functions in the BO routine. One acquisition function of particular interest is Noisy-Input Entropy Search [Fröhlich et al. \(2020\)](#) which searches for robust optima that are less sensitive to noise in their input. Since thermoacoustic stability is highly sensitive to small perturbations in model parameters and geometry, finding more robust optima may lead to better designs.

Acknowledgements

Ushnish Sengupta is an Early Stage Researcher within the MAGISTER consortium which receives funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 766264.

References

- Aguilar JG and Juniper MP (2020) Thermoacoustic stabilization of a longitudinal combustor using adjoint methods. *Physical Review Fluids* 5(8): 083902.
- Aguilar Perez JG (2019) *Sensitivity analysis and optimization in low order thermoacoustic models*. PhD Thesis, University of Cambridge.
- Balachandran R, Ayoola B, Kaminski C, Dowling A and Mastorakos E (2005) Experimental investigation of the nonlinear response of turbulent premixed flames to imposed inlet velocity oscillations. *Combustion and Flame* 143(1-2): 37–55.

- Constantine PG, Dow E and Wang Q (2014) Active subspace methods in theory and practice: applications to kriging surfaces. *SIAM Journal on Scientific Computing* 36(4): A1500–A1524.
- Eriksson D, Dong K, Lee EH, Bindel D and Wilson AG (2018) Scaling gaussian process regression with derivatives. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. pp. 6868–6878.
- Falco S and Juniper MP (2021) Shape optimization of thermoacoustic systems using a two-dimensional adjoint helmholtz solver. *Journal of Engineering for Gas Turbines and Power* 143(7): 071025.
- Fröhlich L, Klenske E, Vinogradska J, Daniel C and Zeilinger M (2020) Noisy-input entropy search for efficient robust bayesian optimization. In: *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 2262–2272.
- Gardner J, Pleiss G, Wu R, Weinberger K and Wilson A (2018) Product kernel interpolation for scalable gaussian processes. In: *International Conference on Artificial Intelligence and Statistics*. PMLR, pp. 1407–1416.
- Huang D, Allen TT, Notz WI and Zeng N (2006) Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of global optimization* 34(3): 441–466.
- Jones DR, Schonlau M and Welch WJ (1998) Efficient global optimization of expensive black-box functions. *Journal of Global optimization* 13(4): 455–492.
- Jones H, Gaudron R and Morgans AS (2021) Shape optimisation of an unstable combustor using a genetic algorithm. In: *Symposium on Thermoacoustics in Combustion: Academia meets Industry*.
- Joos F (2006) Thermoakustik von brennkammern. *Technische Verbrennung: Verbrennungstechnik, Verbrennungsmodellierung, Emissionen* : 737–774.
- Juniper MP (2018) Sensitivity analysis of thermoacoustic instability with adjoint helmholtz solvers. *Physical Review Fluids* 3(11): 110509.
- Lizotte DJ (2008) *Practical bayesian optimization*. University of Alberta.
- Magri L (2019) Adjoint methods as design tools in thermoacoustics. *Applied Mechanics Reviews* 71(2).
- Magri L and Juniper M (2013) Sensitivity analysis of a time-delayed thermo-acoustic system via an adjoint-based approach. *Journal of Fluid Mechanics* 719: 183–202.
- Mensah GA and Moeck JP (2017) Acoustic damper placement and tuning for annular combustors: an adjoint-based optimization study. *Journal of Engineering for Gas Turbines and Power* 139(6).
- Mensah GA, Orchini A and Moeck JP (2019) *Adjoint-based computation of shape sensitivity in a Rijke-tube*. Universitätsbibliothek der RWTH Aachen.
- Scott W, Frazier P and Powell W (2011) The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression. *SIAM Journal on Optimization* 21(3): 996–1026.

-
- Silva CF, Magri L, Runte T and Polifke W (2017) Uncertainty quantification of growth rates of thermoacoustic instability by an adjoint helmholtz solver. *Journal of Engineering for Gas Turbines and Power* 139(1).
- Srinivas N, Krause A, Kakade SM and Seeger M (2009) Gaussian process optimization in the bandit setting: No regret and experimental design. *arXiv preprint arXiv:0912.3995* .
- Wang Z, Zoghi M, Hutter F, Matheson D, De Freitas N et al. (2013) Bayesian optimization in high dimensions via random embeddings. In: *IJCAI*. pp. 1778–1784.
- Williams CK and Rasmussen CE (1996) *Gaussian processes for regression*. MIT.