

A new perspective on parameter study of optimization problems

Alen Alexanderian^a, Joseph Hart^{b,*}, Mason Stevens^a^a Department of Mathematics, North Carolina State University, Raleigh, NC, United States of America^b Department of Scientific Machine Learning, Sandia National Laboratories, Albuquerque, NM, United States of America

ARTICLE INFO

Article history:

Received 23 September 2022

Received in revised form 15 December 2022

Accepted 16 December 2022

Available online 19 December 2022

Keywords:

Optimization

Post-optimal sensitivity analysis

Inverse problems

ABSTRACT

We provide a new perspective on the study of parameterized optimization problems. Our approach combines methods for post-optimal sensitivity analysis and ordinary differential equations to quantify the uncertainty in the minimizer due to uncertain parameters in the optimization problem. We illustrate the proposed approach with a simple analytic example and an inverse problem governed by an advection diffusion equation.

© 2022 Elsevier Ltd. All rights reserved.

1. Introduction

A common class of problems in the sciences and engineering involves solving optimization problems constrained by differential equations. Examples include inverse problems and optimal design or control problems. To illustrate, we consider the following advection diffusion equation:

$$\begin{aligned} -\kappa u'' + vu' &= s & \text{in } (0, 1), \\ \kappa u' &= \alpha u & \text{on } x = 0, \\ \kappa u' &= -\alpha u & \text{on } x = 1. \end{aligned} \quad (1)$$

Here $u(x)$ is the temperature at a point $x \in [0, 1]$, κ is the diffusion coefficient, v is wind velocity, α models a heat transfer coefficient, and s is a source term defined by

$$s(x) = a \exp(-200(x - c)^2), \quad (2)$$

which models a localized source with a and c indicating its magnitude and location.

* Corresponding author.

E-mail address: joshart@sandia.gov (J. Hart).

Suppose we have measurements of the temperature throughout the domain and we seek to use this information to estimate the parameter vector $\mathbf{m} = [\kappa \ v]^\top \in \mathbb{R}^2$. This involves solving an optimization problem of the form,

$$\min_{\mathbf{m}} J(\mathbf{m}) := \frac{1}{2} \int_0^1 (u(x) - u^{\text{obs}}(x))^2 dx + \frac{\beta}{2} \|\mathbf{m} - \mathbf{m}^0\|_2^2, \quad (3)$$

where u is the solution of (1) (which depends on $\mathbf{m}$), u^{obs} is the observed temperature across the domain, $\mathbf{m}^0$ is a prior estimate of $\mathbf{m}$, and $\beta > 0$ is a regularization parameter.

Note that the objective function J also depends on the vector of model parameters $\boldsymbol{\theta} = [a \ c \ \alpha]^\top \in \mathbb{R}^3$ that parameterize the volume source term and the heat transfer. In practice, these parameters might not be known exactly. Thus, it is imperative to understand how the uncertainty in these parameters affects the estimated parameter $\mathbf{m}^*$ obtained by solving (3). In the present work, we propose an approach for analyzing such *parameterized* optimization problems, without the need for repeated solutions of the optimization problem.

The study of parameterized optimization problems can be found in the early works [1,2] followed by more advanced developments in [3–5]. These works, and references therein, provide extensive theory concerning the differentiability of optimal solutions with respect to parameter perturbations. This field has assumed various names as it arises in different parts of the literature. Herein we refer to it as post-optimality sensitivity analysis as it provides a local sensitivity study of the optimal solution. Developments from [6–9] extended the use of post-optimality sensitivities to optimization problems constrained by partial differential equations. Recent work from [10–12] has focused on making this sensitivity analysis scalable for high-dimensional parameter spaces and extending its use for various classes of parametric uncertainty. However, post-optimality sensitivity analysis is local in the sense that it is only valid in a neighborhood of a nominal parameter value. This article borrows concepts from the post-optimal sensitivity analysis literature and couples them with a time stepping algorithm to move through the parameter space to perform efficient global analysis.

We detail the mathematical setup of the parameterized optimization problems under study in Section 2. Our proposed approach is presented in Section 3. We illustrate the effectiveness of our approach in Section 4, for an analytic test problem as well as the inverse advection diffusion problem discussed above. Concluding remarks are given in Section 5.

2. Preliminaries

In this section, we lay out the mathematical setup of the optimization problems under study. Let U be a compact subset of $\mathbb{R}^d$ and consider parameterized optimization problems of the form

$$\min_{\mathbf{m} \in U} J(\mathbf{m}, \boldsymbol{\theta}), \quad (4)$$

where $\boldsymbol{\theta}$ is a vector of parameters. These parameters are fixed when solving the optimization problem, but in practice are uncertain and can be modeled as random variables. We assume that $\boldsymbol{\theta}$ belongs to a compact set $\Theta \subset \mathbb{R}^p$ and $\bar{\boldsymbol{\theta}} \in \Theta$ is a nominal parameter vector. Let $\bar{\mathbf{m}}^*$ be a minimizer of $J(\mathbf{m}, \bar{\boldsymbol{\theta}})$ and $U_0 \subset U$ be an open set that contains $\bar{\mathbf{m}}^*$. To facilitate our parametric study of (4), we assume that

1. for each $\boldsymbol{\theta} \in \Theta$, there exists a unique minimizer $\mathbf{m}^*(\boldsymbol{\theta})$ in U_0 ; and
2. J is twice continuously differentiable in $\mathbf{m}$ and $\boldsymbol{\theta}$.

It follows that $\mathbf{m}^*(\boldsymbol{\theta})$ is a differentiable function on Θ .

Verifying the above assumptions for an optimization problem arising from science or engineering applications would be difficult in general. However, these assumptions have an intuitive interpretation. Namely,

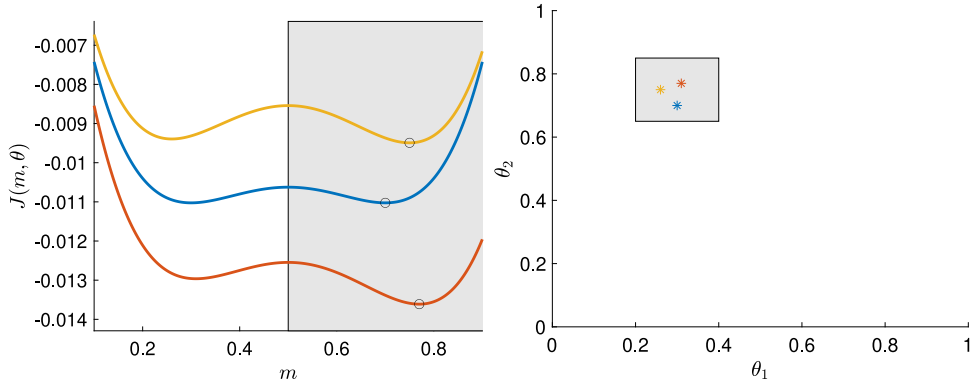


Fig. 1. Illustration of the sets $U_0 = [0.5, 1.0]$ shaded in the left panel and $\Theta = [0.2, 0.4] \times [0.65, 0.85]$ shaded in the right panel. Each curve in the left panel is $J(m, \theta)$, defined by (6), evaluated for a θ sample depicted in the right panel by *. The unique local minimizers in U_0 are denoted by open circles in the left panel.

we consider well-behaved optimization problems whose minimizers are unique in a suitable region around nominal parameters $\bar{\theta}$ and minimizer $\bar{m}^*$. Such assumptions are reasonable in many optimization problems arising in physical applications. Also, in practice, the set Θ will be a user-specified region around the nominal parameter vector $\bar{\theta}$. Specifically, we consider the typical situation where Θ is defined as

$$\Theta = [\bar{\theta}_1 - \varepsilon_1, \bar{\theta}_1 + \varepsilon_1] \times [\bar{\theta}_2 - \varepsilon_2, \bar{\theta}_2 + \varepsilon_2] \times \cdots \times [\bar{\theta}_p - \varepsilon_p, \bar{\theta}_p + \varepsilon_p], \quad (5)$$

where ε_k is some percentage of the corresponding nominal value $\bar{\theta}_k$, $k = 1, \dots, p$. The ε_k 's indicate the level of uncertainty in the physical parameters in the model. It is common that the parameters θ_k are assumed to be uniformly distributed random variables on the respective intervals, but more general distributions on compact sets are admissible.

To illustrate these concepts, consider the objective function

$$J(m, \theta) = \int (m - \theta_1)(m - 0.5)(m - \theta_2) dm, \quad (6)$$

which, for $\theta \in \{\theta \in \mathbb{R}^2 : 0 < \theta_1 < 0.5 \text{ and } 0.5 < \theta_2 < 1\}$, will have two local minima at θ_1 and θ_2 . Letting $U = [0, 1]$ and $\Theta = [0.2, 0.4] \times [0.65, 0.85]$, we may take $U_0 = (0.5, 1.0)$ to restrict our analysis to the local minima at θ_2 . This is depicted in Fig. 1 to demonstrate how the choice of U_0 ensures minimizer uniqueness needed to enable our parameter study.

3. Method

In this section, we outline an approach for approximating $\mathbf{m}^*(\tilde{\theta})$, where $\tilde{\theta}$ is a generic element of Θ .

3.1. An initial value problem for $\mathbf{m}^*(\tilde{\theta})$

To study how the optimal solution changes with θ we compute the Jacobian of $\mathbf{m}^*$ with respect to θ . This Jacobian, denoted by $\mathbf{D}$, can be computed by differentiating through first order optimality condition

$$\frac{\partial J}{\partial m}(\mathbf{m}^*(\theta), \theta) = 0,$$

implicitly. It follows from the Implicit Function Theorem that

$$\mathbf{D}(\mathbf{m}^*(\theta), \theta) = -\mathbf{H}(\mathbf{m}^*(\theta), \theta)^{-1} \mathbf{B}(\mathbf{m}^*(\theta), \theta),$$

where

$$\mathbf{H}(\mathbf{m}^*(\boldsymbol{\theta}), \boldsymbol{\theta}) = \frac{\partial^2 J}{\partial \mathbf{m}^2} \Big|_{(\mathbf{m}^*(\boldsymbol{\theta}), \boldsymbol{\theta})} \quad \text{and} \quad \mathbf{B}(\mathbf{m}^*(\boldsymbol{\theta}), \boldsymbol{\theta}) = \frac{\partial^2 J}{\partial \mathbf{m} \partial \boldsymbol{\theta}} \Big|_{(\mathbf{m}^*(\boldsymbol{\theta}), \boldsymbol{\theta})}.$$

The Jacobian $\mathbf{D}$ is known as the post-optimality sensitivity operator.

For a given $\tilde{\boldsymbol{\theta}} \in \Theta$, we begin by considering the points on the line-segment joining $\bar{\boldsymbol{\theta}}$ to $\tilde{\boldsymbol{\theta}}$,

$$\boldsymbol{\theta}(t) = \bar{\boldsymbol{\theta}} + t(\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}), \quad t \in [0, 1]. \quad (7)$$

Based on our assumptions on J and $\mathbf{m}^*$, we can consider $\mathbf{m}^*$ as a differentiable function of t , $\mathbf{m}^*(t) \equiv \mathbf{m}^*(\boldsymbol{\theta}(t))$. Taking the derivative of $\mathbf{m}^*$ with respect to t and applying the Chain Rule gives

$$\frac{d\mathbf{m}^*}{dt} = \frac{\partial \mathbf{m}^*}{\partial \boldsymbol{\theta}} \cdot \frac{d\boldsymbol{\theta}}{dt} = \mathbf{D}(\mathbf{m}^*(\boldsymbol{\theta}(t)), \boldsymbol{\theta}(t))(\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}). \quad (8)$$

For notational convenience, we define

$$\mathbf{f}(t, \mathbf{m}^*) = -\mathbf{H}(\mathbf{m}^*, \boldsymbol{\theta}(t))^{-1} \mathbf{B}(\mathbf{m}^*, \boldsymbol{\theta}(t))(\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}). \quad (9)$$

Then we may determine $\mathbf{m}^*(t)$, for $t \in [0, 1]$, by solving the following initial value problem (IVP)

$$\begin{aligned} \frac{d\mathbf{m}^*}{dt} &= \mathbf{f}(t, \mathbf{m}^*), \\ \mathbf{m}^*(0) &= \mathbf{m}^*(\bar{\boldsymbol{\theta}}). \end{aligned} \quad (10)$$

To specify the initial condition, the optimization problem (4) needs to be solved. This is the only solution of the optimization problem required in our approach. For each parameter $\tilde{\boldsymbol{\theta}}$, we will solve the IVP (10) up to $t = 1$ to determine the corresponding local minimum $\mathbf{m}^*(1) \equiv \mathbf{m}^*(\boldsymbol{\theta}(1)) = \mathbf{m}^*(\tilde{\boldsymbol{\theta}})$. The right hand side function $\mathbf{f}$ (9) is the post optimality sensitivity operator acting on $\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}$ and hence the IVP depends on the parameter sample $\tilde{\boldsymbol{\theta}}$.

3.2. Time-stepping to approximate $\mathbf{m}^*(\tilde{\boldsymbol{\theta}})$

We can apply common numerical methods to solve the IVP (10). In this work, we study the use of forward Euler. Specifically, let $h = 1/N$ be a step-size and $t_n = nh$, $n = 0, \dots, N$. The forward Euler discretization of (10) is

$$\mathbf{m}_{n+1}^* = \mathbf{m}_n^* + h \mathbf{f}(t_n, \mathbf{m}_n^*), \quad n = 0, 1, \dots, N-1, \quad (11)$$

where $\mathbf{m}_0^* = \mathbf{m}^*(0)$ and $\mathbf{m}_n^* \approx \mathbf{m}^*(t_n)$, $n = 1, \dots, N$. Finally, the approximation to $\mathbf{m}^*(\tilde{\boldsymbol{\theta}})$ is given by

$$\mathbf{m}^*(\tilde{\boldsymbol{\theta}}) \equiv \mathbf{m}^*(t_N) \approx \mathbf{m}_N^*. \quad (12)$$

Notice that (11) resembles Newton's method for optimization. In particular, computing $\mathbf{m}^*(\tilde{\boldsymbol{\theta}})$ via Newton's method iterates with search directions of the form

$$-\mathbf{H}(\mathbf{m}_n^{\text{Newton}}, \tilde{\boldsymbol{\theta}})^{-1} \mathbf{g}(\mathbf{m}_n^{\text{Newton}}, \tilde{\boldsymbol{\theta}}),$$

where $\mathbf{m}_n^{\text{Newton}}$ denotes the n th Newton iterate and $\mathbf{g}(\mathbf{m}_n^{\text{Newton}}, \tilde{\boldsymbol{\theta}})$ denotes the gradient of J with respect to $\mathbf{m}$, evaluated at $\mathbf{m}_n^{\text{Newton}}$. On the other hand, time marching via (11) has search directions of the form

$$-\mathbf{H}(\mathbf{m}_n^*, \boldsymbol{\theta}(t_n))^{-1} \mathbf{B}(\mathbf{m}_n^*, \boldsymbol{\theta}(t_n))h(\tilde{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}) \approx -\mathbf{H}(\mathbf{m}_n^*, \boldsymbol{\theta}(t_n))^{-1} (\mathbf{g}(\mathbf{m}_n^*, \boldsymbol{\theta}(t_{n+1})) - \mathbf{g}(\mathbf{m}_n^*, \boldsymbol{\theta}(t_n))),$$

where the latter approximation follows since $\mathbf{B}$ is the derivative of $\mathbf{g}$ with respect to $\boldsymbol{\theta}$. Note that $\mathbf{g}(\mathbf{m}_n^*, \boldsymbol{\theta}(t_n))$ is expected to be small since it is the gradient evaluated at an approximate minimizer. Time

marching via (11) takes a size h perturbation of θ and updates the solution via a Newton like step whereas resolving the optimization problem takes the full parameter step $\tilde{\theta} - \hat{\theta}$ and then employs Newton iterations to update $\mathbf{m}$. Note also that in resolving the optimization problem via Newton's method, employing a line search algorithm such as Armijo's method [13] is typically necessary.

An alternate point of view regarding the time stepping (11) is performing continuation on the parameters $\theta_0, \theta_1, \dots, \theta_N$ and using an approximate Newton step to update $\mathbf{m}$ after each parameter perturbation.

A benefit of solving (10) is that computing $\mathbf{f}(t_n, \mathbf{m}_n^*) = \mathbf{D}(\mathbf{m}_n^*, \theta_n)(\tilde{\theta} - \bar{\theta})$ at each time step provides post-optimality sensitivity information in the direction $\tilde{\theta} - \bar{\theta}$. This gives additional insight about how the minimizer depends on the uncertain parameters.

3.3. Computational considerations

Computing (9) requires access to the Hessian $\mathbf{H}$ and the matrix $\mathbf{B}$ of mixed second order partial derivatives. For optimization problems governed by partial differential equations, such derivative information can be obtained efficiently using adjoint state methods; see e.g., [14]. Specifically, in that context, one obtains adjoint based expressions for applying $\mathbf{H}$ and $\mathbf{B}$ to vectors [11]. In large-scale computations, the inverse Hessian apply is computed by performing a linear solve using the Conjugate-Gradient method, which only requires applications of the Hessian on vectors. An alternative approach for obtaining the required derivatives is automatic differentiation. Simple finite-difference approaches might be applicable as well if the gradient can be computed exactly and differenced to approximate the second derivatives.

It is also possible to use methods other than forward Euler to numerically solve the IVP (10). Generally, implicit methods would be very challenging to implement for the present IVP, due the requirement of a nonlinear solve in each step. On the other hand, higher order explicit Runge-Kutta methods or predictor corrector methods will be straightforward to implement. However, we caution that the faster convergence might come at a cost of making the time-stepping more expensive than resolving the optimization problem for different realizations of $\tilde{\theta}$. A thorough investigation of the time-stepping approaches that are tractable for (10) and analysis of the related computational cost is beyond scope of the present study and will be pursued in future work.

4. Numerical examples

4.1. A one-dimensional example

Consider the function

$$J(m, \theta) = \frac{\theta_1}{1 + e^{\theta_2 m}} + \theta_3 m^2, \quad m \in \mathbb{R}, \theta \in \Theta.$$

Here $\Theta \subset \mathbb{R}^3$ corresponds to taking 40% of θ around their nominal values $\bar{\theta} = [1 \quad 3 \quad 0.1]^\top$. In Fig. 2 (left), we show J with $\theta = \bar{\theta}$, and in Fig. 2 (right), we display several realizations of the function $J(m, \theta)$, corresponding to random draws from the (uniform) distribution of θ . This demonstrates significant variations in the location of the minimizer.

In Fig. 3 (left), we show the probability density function (pdf) of $\mathbf{m}_N^*$, for a few choices of N in (11). We also track the convergence of the mean and standard deviation in Fig. 3 (middle/right). The optimization problem was solved for 5000 realizations of θ to generate a reference distribution for $\mathbf{m}^*(\theta)$. From Fig. 3 we see that a small N is sufficient for approximating the pdf of $\mathbf{m}^*$. In fact, $N = 1$ provides a reasonable approximation, and as N grows, the pdf of $\mathbf{m}_N^*$ approaches that of $\mathbf{m}^*$ rapidly. Also, the mean and standard deviation exhibit a first order convergence consistent with the convergence rate of forward Euler.

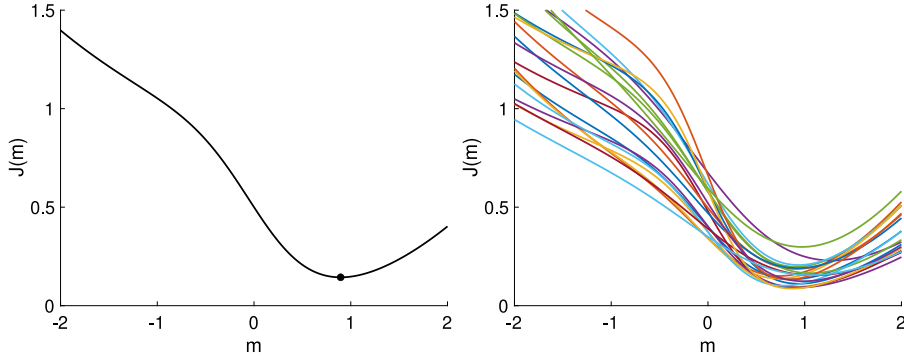


Fig. 2. The nominal model $J(m, \bar{\theta})$ (left) and several realizations of the model (right) for the one-dimensional example. In the left panel, the black dot indicates the location of the minimizer.

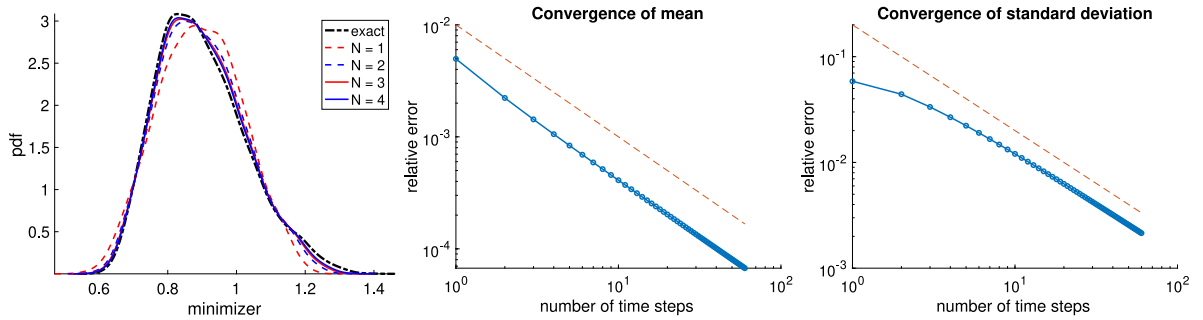


Fig. 3. Convergence of the pdfs (left), the mean (middle), and standard deviation (right), as N increases, for the one-dimensional example. In the middle and right panels, the dashed lines indicate $\mathcal{O}(h)$, where $h = 1/N$.

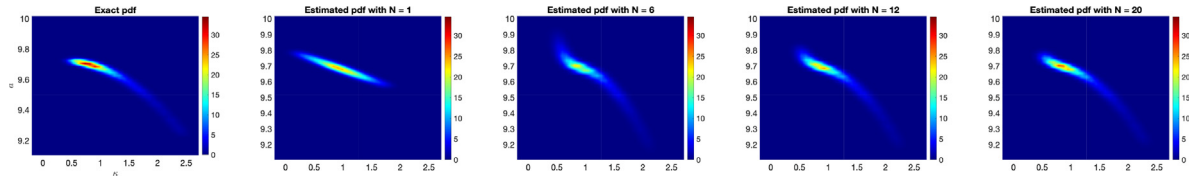


Fig. 4. Convergence of the joint pdfs for increasing N on the differential equation constrained example.

4.2. A differential equation constrained example

We revisit our illustrative optimization problem (3) and consider estimating the pdf of the optimal solution. We draw 5000 parameter samples from a uniform distribution modeling $\pm 20\%$ uncertainty around the nominal parameter vector $\bar{\theta} = [10.0 \ 0.05 \ 1.0]^\top$. Fig. 4 displays the joint pdf computed by solving the optimization problem for each parameter sample (left) and compares it with the estimated joint pdf coming from our proposed approach using $N = 1, 6, 12$, and 20 time steps. In each of these cases, we solve the optimization problem once for $\theta = \bar{\theta}$ and then solve the IVP (10) for each parameter sample to estimate the minimizer. Using the same samples, Fig. 5 shows the convergence of marginal pdfs of the minimizer. We observe that some information about the correlation structure in the joint pdf and the marginal pdf for κ can be inferred with a small N . The marginal pdf of a exhibits complex features that are not easily resolved with a small N ; nonetheless, for modest values of N we are able to capture many of its features.

We next compare the computational cost of our proposed time-stepping approach against repeatedly solving the optimization problem for each realization of θ . We use a trust region algorithm for solving the optimization problem, whose cost per iteration is approximately the same as that of a time step (11). In

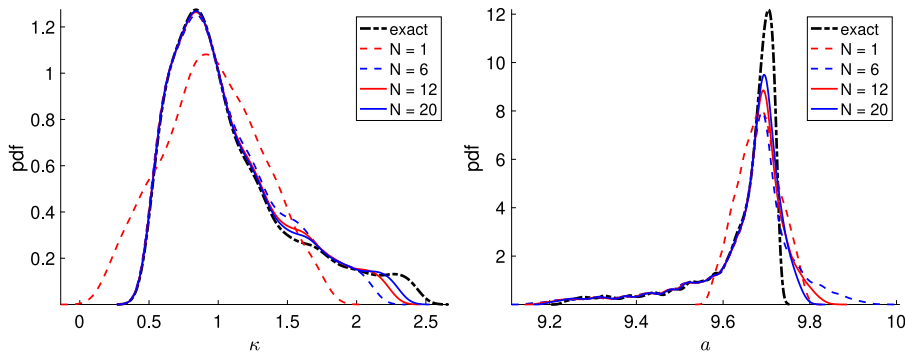


Fig. 5. Convergence of the marginal pdfs for increasing N on the differential equation constrained example.

the present study, we average the number of iterations required by the optimization algorithm when solving the optimization problem for each of the 5000 parameter samples. We observed an average of 6.58 iterations when resolving the optimization problem. This cost is slightly greater than that of using $N = 6$ time steps to approximate the minimizer via our proposed approach. However, there are many nuisances related to optimization algorithm and algorithmic parameter choices. Our comparison here highlights how the cost are comparable in a simplified setting but a more detailed study is required in future work.

5. Conclusion

The time-stepping approach introduced in this article is a new perspective on a classical problem of studying parametric uncertainty in optimization problems. Tools from post-optimality sensitivity analysis have traditionally been used to perform local parameter studies [4]. By formulating an ordinary differential equation driven by the post-optimality sensitivity, our approach offers a mathematically rigorous approach to transition from local to global parameter studies. We conjecture that with suitable time discretizations, the computational cost of our approach can be less than the cost of resolving the optimization problem for each new parameter sample. Many questions remain. These include the trade-offs between higher order time integration schemes, finer temporal discretizations, and stability of the time stepping. Further, there may be opportunities to leverage information from previous time steps for preconditioning of future solves or reusing time steps to explore multiple parameter samples. Exploiting such structure may enable further gains in the computational performance of our approach relative to the base line of repeatedly resolving optimization problems.

Optimization problems constrained by partial differential equations (PDEs) frequently translate to high-dimensional parameters that result from the discretization of functions. The proposed approach will scale efficiently to such problems. Using the adjoint method for derivative calculation, the action of the Hessian or the mixed second derivative matrix on a vector can be computed using two linear PDE solves, regardless of the parameter dimension. The curse of dimensionality appears in the potential need for many matrix–vector products to explore the space; however, a low-rank structure is commonly present and can be exploited to accelerate computations.

Another area of inquiry is to extract global sensitivity information alongside the distribution of the optimal solution. Each time step computes the action of the post-optimality sensitivity operator for a different sample. Understanding how to aggregate this sensitivity information will provide valuable insights that are not available from repeated optimization solves. Such global aggregation of sensitivity information is common in the derivative-based global sensitivity analysis literature [15,16] and uncertainty quantification more broadly [17].

Data availability

No data was used for the research described in the article.

Acknowledgments

This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan <https://www.energy.gov/downloads/doe-public-access-plan>.

Sandia National Laboratories is a multitechnology laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. SAND2023-10468J.

This work was supported by the US Department of Energy, Office of Advanced Scientific Computing Research, Field Work Proposal 20-023231.

The work of A. Alexanderian and M. Stevens was supported in part by National Science Foundation, United States of America under grant DMS-1745654. The work of A. Alexanderian was also supported in part by the National Science Foundation, United States of America under grant DMS-2111044.

References

- [1] A.V. Fiacco, G.P. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York, 1968.
- [2] A.V. Fiacco, Sensitivity analysis for nonlinear programming using penalty methods, *Math. Program.* 10 (3) (1976) 287–311.
- [3] J.F. Bonnans, Directional derivatives of optimal solutions in smooth nonlinear programming, *J. Optim. Theory Appl.* 73 (1) (1992) 27–45.
- [4] J.F. Bonnans, A. Shapiro, Optimization problems with perturbations: A guided tour, *SIAM Rev.* 40 (2) (1998) 228–264.
- [5] J.F. Bonnans, A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer, 2013.
- [6] R. Griesse, Parametric sensitivity analysis in optimal control of a reaction diffusion system. I. Solution differentiability, *Numer. Funct. Anal. Optim.* 25 (1–2) (2004) 93–117.
- [7] R. Griesse, Parametric sensitivity analysis in optimal control of a reaction-diffusion system – part II: practical methods and examples, *Optim. Methods Softw.* 19 (2) (2004) 217–242.
- [8] R. Griesse, *Stability and Sensitivity Analysis in Optimal Control of Partial Differential Equations* (Habilitation Thesis), Karl-Franzens University Graz, 2007.
- [9] K. Brandes, R. Griesse, Quantitative stability analysis of optimal solutions in PDE-constrained optimization, *J. Comput. Appl. Math.* (2006).
- [10] J. Hart, B. van Bloemen Waanders, R. Herzog, Hyper-differential sensitivity analysis of uncertain parameters in PDE-constrained optimization, *Int. J. Uncertain. Quantif.* 10 (3) (2020) 225–248.
- [11] I. Sunseri, J. Hart, B. van Bloemen Waanders, A. Alexanderian, Hyper-differential sensitivity analysis for inverse problems constrained by partial differential equations, *Inverse Problems* 36 (12) (2020) 125001, <http://dx.doi.org/10.1088/1361-6420/abaf63>.
- [12] A.K. Saibaba, J. Hart, B. van Bloemen Waanders, Randomized algorithms for generalized singular value decomposition with application to sensitivity analysis, *Numer. Linear Algebra Appl.* 28 (4) (2021).
- [13] S. Wright, J. Nocedal, *Numerical optimization*, Springer Science 35 (67–68) (1999) 7.
- [14] M.D. Gunzburger, *Perspectives in Flow Control and Optimization*, SIAM, 2003.
- [15] I. Sobol', S. Kucherenko, Derivative based global sensitivity measures and the link with global sensitivity indices, *Math. Comput. Simulation* 79 (2009) 3009–30017.

- [16] I. Sobol', S. Kucherenko, A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices, *Comput. Phys. Comm.* 181 (2010) 1212–1217.
- [17] R.C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications*, in: *Computational Science and Engineering Series*, vol. 12, SIAM, 2013.