Second-Order Corrections for Surrogate-Based Optimization with Model Hierarchies

M. S. Eldred[*], A. A. Giunta[†], and S. S. Collis[*]
Sandia National Laboratories[‡], Albuquerque, NM 87185

Surrogate-based optimization methods have become established as effective techniques for engineering design problems through their ability to tame nonsmoothness and reduce computational expense. In recent years, supporting mathematical theory has been developed to provide the foundation of provable convergence for these methods. One of the requirements of this provable convergence theory involves consistency between the surrogate model and the underlying truth model that it approximates. This consistency can be enforced through a variety of correction approaches, and is particularly essential in the case of surrogate-based optimization with model hierarchies. First-order additive and multiplicative corrections currently exist which satisfy consistency in values and gradients between the truth and surrogate models at a single point. This paper demonstrates that first-order consistency can be insufficient to achieve acceptable convergence rates in practice and presents new second-order additive, multiplicative, and combined corrections which can significantly accelerate convergence. These second-order corrections may enforce consistency with either the actual truth model Hessian or its finite difference, quasi-Newton, or Gauss-Newton approximation.

Introduction

OPTIMIZATION methods employing approximation models originated in the 1970's[1] and have proved extremely popular within the engineering community. Numerous surveys of these methods exist.[2,3] However, many of these methods have been inherently heuristic, lacking the mathematical rigor necessary to have predictable performance. In particular, they performed well on some problems, and failed to converge to a minimum of the original model on others.

In recent years, supporting mathematical theory has been developed to provide the foundation of provable convergence for a broad class of approximation-based optimization methods. The terms "surrogate-based optimization" and "model management framework" are used to describe these rigorous methods.[4,5] Provided that one employs a sufficiently rigorous globalization approach (e.g., trust region management) and satisfies first-order consistency between the surrogate model and the underlying truth model, then convergence of the surrogate-based optimization process to the optimum of the high-fidelity model can be guaranteed (note: this discussion focuses on first-order model management; provable convergence theory also exists for zeroth-order model management based on pattern search, but will not be discussed here).

A number of surrogate model selections are possible. First, the surrogate may be of the "data fit" type, which is a non-physics-based approximation typically involving interpolation or regression of a set of data generated from the high-fidelity model. Data fit surrogates can be further characterized by the number of data points used in the fit, where a local approximation (e.g., Taylor series) uses data from a single point, a multipoint approximation (e.g., two-point adaptive nonlinear approximations) uses a small number of data points often drawn from the previous iterates of a particular algorithm, and a global approximation (e.g., polynomial response surfaces, kriging, neural networks, radial basis functions, splines) uses a set of data points distributed over the domain of interest (often generated using a design of computer experiments). A completely different type of surrogate is the model hierarchy type (also called multifidelity, variable fidelity, variable complexity, etc.). In this case, a model that is still physics-based but is of lower fidelity (e.g., coarser discretization, looser convergence tolerances, omitted physics) is used as the surrogate in place of the high-fidelity model. Reduced-order modeling techniques such as proper orthogonal decomposition in computational fluid dynamics or modal analysis in structural dynamics are also considered to be in the model hierarchy family due to their physics linkages.

When performing surrogate-based optimization with local/multipoint/global data fit surrogates, it is necessary to regenerate or update the data fit for each new trust region. In the global data fit case, this can mean performing a new design of experiments on the original high-fidelity model for each trust region, which

[*]Optimization and Uncertainty Estimation Department, Senior Member AIAA.
[†]Validation and Uncertainty Quantification Processes Department, Senior Member AIAA.
[‡]Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed-Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000

can effectively limit the approach to use on problems with, at most, tens of variables. However, an important benefit of the global sampling is that the global data fits can tame poorly-behaved, nonsmooth, discontinuous response variations within the high-fidelity model into smooth, differentiable, easily navigated surrogates. When enforcing local consistency between a global data fit surrogate and a truth model at a point, care must be taken to balance this local consistency requirement with the global accuracy of the surrogate. In particular, performing a correction on an existing global data fit in order to enforce local consistency can skew the data fit and destroy its global accuracy. A better approach is to include the consistency requirement within the data fit process by constraining the global data fit calculation (e.g., using constrained least squares). This allows the data fit to satisfy the consistency requirement while still addressing global accuracy with its remaining degrees of freedom. Alternatively, one can employ a surrogate that directly enforces local consistency through its assumed form, e.g., a second-order Taylor series with globally estimated Hessian terms[6] will inherently satisfy first-order consistency.

When performing surrogate-based optimization with model hierarchies, the low-fidelity model is normally fixed, requiring only a single high-fidelity evaluation to compute a new correction for each new trust region (exception: one may wish to also update the basis vectors in reduced-order modeling methods). This renders the multifidelity surrogate-based optimization technique more scalable to larger numbers of design variables since the number of high-fidelity evaluations per iteration (assuming no finite differencing for derivatives) is independent of the scale of the design problem. However, the ability to smooth poorly-behaved response variations is lost, and the technique becomes dependent on having a well-behaved low-fidelity model. When applying corrections to the low fidelity model, there is no global accuracy concern to balance with the local consistency requirements. Rather, there is only a single truth model evaluation at the center of each trust region, and it is critical to use the best correction possible on the low-fidelity model in order to achieve rapid convergence rates to the optimum of the high-fidelity truth model. The remainder of the paper is focused on this issue.

Correction approaches are closely related to data fit surrogates. As with data fit surrogates, correction approaches may be local, multipoint, or global. As will be shown in Eqs. 8-9, local corrections are derived by generating Taylor series approximations to the ratio or difference between low and high-fidelity models. A multipoint correction will be shown in Eqs. 29-30, in which additive and multiplicative local corrections are combined in such a way as to satisfy an additional matching condition at a previous design iterate. Finally, global corrections use global data fit surrogates to model the relationship (difference or ratio) between low and high-fidelity models at distributed sets of points. A benefit to this latter approach is that the relationship between two model fidelities can tend to be more linear or well-behaved than the models themselves. However, the consistency enforcement with global correction approaches is often zeroth-order (e.g., kriging) or worse (e.g., polynomial regression), which falls short of satisfying the requirements of the provable convergence theory. For each of these correction approaches, the correction is applied to the low-fidelity model to create an approximation to the original high-fidelity model which is then interfaced with the optimization algorithm for search over the current trust region.

The simplest correction approaches are those that enforce consistency in function values between the low and high-fidelity models at a single point in parameter space through use of a simple scalar offset or scaling applied to the low-fidelity model. These zeroth-order approaches cannot change the location of an unconstrained minimum for the low-fidelity model ($\nabla f_{lo}(\mathbf{x})$ will vanish at the same $\mathbf{x}$), and therefore are very limited in their ability to morph the low-fidelity model into providing a good surrogate for minimization of the high-fidelity model. First-order corrections such as the first-order multiplicative correction (also known as beta correction[7]) and the first-order additive correction[8] provide a much more substantial correction capability and are sufficient for ensuring provable convergence of the algorithm. However, the convergence rates can be similar to those achieved by first-order optimization methods such as steepest-descent or sequential linear programming. More successful optimization methods (such as sequential quadratic programming) use at least approximate second-order information to achieve super-linear or quadratic convergence rates in the neighborhood of the minimum, and one would expect the same principle to hold for correction approaches within surrogate-based optimization methods.

In the following sections, important features of the surrogate-based optimization algorithm are described, followed by derivation and discussion of surrogate correction approaches. These techniques are then applied to several computational experiments and concluding remarks are presented.

Surrogate-Based Optimization

Starting from a nonlinear programming problem of the form

$$\begin{aligned} &\text{minimize} && f(\mathbf{x}) \\ &\text{subject to} && \mathbf{x}_l \leq \mathbf{x} \leq \mathbf{x}_u, \end{aligned} \tag{1}$$

where $\mathbf{x} \in \Re^n$ is the vector of design variables, and $\mathbf{x}_l$ and $\mathbf{x}_u$ are the lower and upper bounds, respec-

tively, the surrogate-based optimization (SBO) algorithm solves a sequence of k trust region optimization subproblems of the form

$$\begin{aligned} \text{minimize} \quad & \hat{f}^k(\mathbf{x}) \\ \text{subject to} \quad & \mathbf{x}_l \le \mathbf{x} \le \mathbf{x}_u \\ & \| \mathbf{x} - \mathbf{x}_c^k \|_\infty \le \Delta^k, \end{aligned} \tag{2}$$

where the surrogate model is denoted as $\hat{f}(\mathbf{x})$, $\mathbf{x}_c$ is the center point of the trust region, and the initial value for Δ at $k = 0$ is user-selected. In the multifidelity context, solution of this subproblem involves optimizing the corrected low-fidelity model over the range of the trust region. After each of the k iterations in the SBO strategy, the predicted step is validated by computing $f(\mathbf{x}_*)$ and the trust region ratio ρ is calculated as

$$\rho^k = \frac{f(\mathbf{x}_c^k) - f(\mathbf{x}_*^k)}{\hat{f}(\mathbf{x}_c^k) - \hat{f}(\mathbf{x}_*^k)}, \tag{3}$$

which is the ratio of the actual improvement to the improvement predicted by optimization on the surrogate model. In the multifidelity context, this ratio measures the performance of the corrected low-fidelity model in finding new iterates that improve the high-fidelity objective. The value for ρ then defines the step acceptance and the next trust region size Δ^k using logic similar to:

1. $\rho^k \le 0$ – The surrogates are inaccurate. Reject the step and shrink the trust region by half to improve surrogate accuracy.

2. $0 < \rho^k \le 0.25$ – The surrogates are marginally accurate. Accept the step but shrink the trust region size by half.

3. $0.25 < \rho^k < 0.75$ or $\rho^k > 1.25$ – The surrogates are moderately accurate. Accept the step and maintain the current trust region size.

4. $0.75 \le \rho^k \le 1.25$ – The surrogates are accurate. Accept the step and increase the trust region size by a factor of two.

For simplicity of presentation of the relevant SBO concepts, we have omitted nonlinear inequality and equality constraints $g_i(\mathbf{x}) \le 0$ and $h_i(\mathbf{x}) = 0$ and their surrogates $\hat{g}_i(\mathbf{x})$ and $\hat{h}_i(\mathbf{x})$. Constraint management in SBO can be approached in several different ways. One approach is to use an augmented Lagrangian function to combine the objective and nonlinear constraints.[9,10] Penalty-free methods based on filter methods are also under investigation which manage constraint relaxations within the trust regions using homotopy or fraction of Cauchy decrease approaches.[11]

Surrogate Corrections

Computing local corrections

A variety of relationships between the high-fidelity and low-fidelity models can be used in deriving correction approaches. The primary two of interest are

$$A(\mathbf{x}) = f_{hi}(\mathbf{x}) - f_{lo}(\mathbf{x}) \tag{4}$$

$$B(\mathbf{x}) = \frac{f_{hi}(\mathbf{x})}{f_{lo}(\mathbf{x})} \tag{5}$$

which correspond to the (exact) additive correction

$$f_{hi}(\mathbf{x}) = f_{lo}(\mathbf{x}) + A(\mathbf{x}) \tag{6}$$

and to the (exact) multiplicative correction

$$f_{hi}(\mathbf{x}) = f_{lo}(\mathbf{x}) B(\mathbf{x}) \tag{7}$$

We will not be computing the exact correction functions, but rather approximating $A(\mathbf{x})$ and $B(\mathbf{x})$ as $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$. This approximation may involve a local, multipoint, or global approximation. For the case of a local approximation, consider the second-order Taylor series expansions centered at $\mathbf{x_c}$:

$$\begin{aligned} \alpha(\mathbf{x}) = {} & A(\mathbf{x_c}) + \nabla A(\mathbf{x_c})^T (\mathbf{x} - \mathbf{x_c}) + \\ & \frac{1}{2} (\mathbf{x} - \mathbf{x_c})^T \nabla^2 A(\mathbf{x_c}) (\mathbf{x} - \mathbf{x_c}) \end{aligned} \tag{8}$$

$$\begin{aligned} \beta(\mathbf{x}) = {} & B(\mathbf{x_c}) + \nabla B(\mathbf{x_c})^T (\mathbf{x} - \mathbf{x_c}) + \\ & \frac{1}{2} (\mathbf{x} - \mathbf{x_c})^T \nabla^2 B(\mathbf{x_c}) (\mathbf{x} - \mathbf{x_c}) \end{aligned} \tag{9}$$

where, by differentiating Eq. 4 twice, it can be shown that

$$A(\mathbf{x_c}) = f_{hi}(\mathbf{x_c}) - f_{lo}(\mathbf{x_c}) \tag{10}$$

$$\nabla A(\mathbf{x_c}) = \nabla f_{hi}(\mathbf{x_c}) - \nabla f_{lo}(\mathbf{x_c}) \tag{11}$$

$$\nabla^2 A(\mathbf{x_c}) = \nabla^2 f_{hi}(\mathbf{x_c}) - \nabla^2 f_{lo}(\mathbf{x_c}) \tag{12}$$

and, by differentiating Eq. 5 twice, it can be shown that

$$B(\mathbf{x_c}) = \frac{f_{hi}(\mathbf{x_c})}{f_{lo}(\mathbf{x_c})} \tag{13}$$

$$\nabla B(\mathbf{x_c}) = \frac{1}{f_{lo}(\mathbf{x_c})} \nabla f_{hi}(\mathbf{x_c}) - \frac{f_{hi}(\mathbf{x_c})}{f_{lo}^2(\mathbf{x_c})} \nabla f_{lo}(\mathbf{x_c}) \tag{14}$$

$$\begin{aligned} \nabla^2 B(\mathbf{x_c}) = {} & \frac{1}{f_{lo}(\mathbf{x_c})} \nabla^2 f_{hi}(\mathbf{x_c}) - \frac{f_{hi}(\mathbf{x_c})}{f_{lo}^2(\mathbf{x_c})} \nabla^2 f_{lo}(\mathbf{x_c}) + \\ & \frac{2 f_{hi}(\mathbf{x_c})}{f_{lo}^3(\mathbf{x_c})} \nabla f_{lo}(\mathbf{x_c}) \nabla f_{lo}^T(\mathbf{x_c}) - \\ & \frac{1}{f_{lo}^2(\mathbf{x_c})} (\nabla f_{lo}(\mathbf{x_c}) \nabla f_{hi}^T(\mathbf{x_c}) + \\ & \nabla f_{hi}(\mathbf{x_c}) \nabla f_{lo}^T(\mathbf{x_c})) \end{aligned} \tag{15}$$

Additional correction formulations are easily derived using this approach. For example, the following subtractive and division-based corrections

$$f_{hi}(\mathbf{x}) = f_{lo}(\mathbf{x}) - C(\mathbf{x}) \tag{16}$$

$$f_{hi}(\mathbf{x}) = \frac{f_{lo}(\mathbf{x})}{D(\mathbf{x})} \tag{17}$$

could be motivated in some instances and would involve a simple swapping of high and low values in Eqs. 10-15.

Applying local corrections

Correcting low-fidelity function values, gradients, and Hessians with additive corrections involves

$$\hat{f_{hi_\alpha}}(\mathbf{x}) = f_{lo}(\mathbf{x}) + \alpha(\mathbf{x}) \quad (18)$$
$$\nabla\hat{f_{hi_\alpha}}(\mathbf{x}) = \nabla f_{lo}(\mathbf{x}) + \nabla\alpha(\mathbf{x}) \quad (19)$$
$$\nabla^2\hat{f_{hi_\alpha}}(\mathbf{x}) = \nabla^2 f_{lo}(\mathbf{x}) + \nabla^2\alpha(\mathbf{x}) \quad (20)$$

and with multiplicative corrections involves

$$\hat{f_{hi_\beta}}(\mathbf{x}) = f_{lo}(\mathbf{x})\beta(\mathbf{x}) \quad (21)$$
$$\nabla\hat{f_{hi_\beta}}(\mathbf{x}) = f_{lo}(\mathbf{x})\nabla\beta(\mathbf{x}) + \beta(\mathbf{x})\nabla f_{lo}(\mathbf{x}) \quad (22)$$
$$\nabla^2\hat{f_{hi_\beta}}(\mathbf{x}) = f_{lo}(\mathbf{x})\nabla^2\beta(\mathbf{x}) + \nabla\beta(\mathbf{x})\nabla f_{lo}^T(\mathbf{x}) + \nabla f_{lo}(\mathbf{x})\nabla\beta^T(\mathbf{x}) + \beta(\mathbf{x})\nabla^2 f_{lo}(\mathbf{x}) \quad (23)$$

where $\alpha(\mathbf{x})$ and $\beta(\mathbf{x})$ are as defined in Eqs. 8 and 9 and

$$\nabla\alpha(\mathbf{x}) = \nabla A(\mathbf{x_c}) + \nabla^2 A(\mathbf{x_c})(\mathbf{x} - \mathbf{x_c}) \quad (24)$$
$$\nabla\beta(\mathbf{x}) = \nabla B(\mathbf{x_c}) + \nabla^2 B(\mathbf{x_c})(\mathbf{x} - \mathbf{x_c}) \quad (25)$$
$$\nabla^2\alpha(\mathbf{x}) = \nabla^2 A(\mathbf{x_c}) \quad (26)$$
$$\nabla^2\beta(\mathbf{x}) = \nabla^2 B(\mathbf{x_c}) \quad (27)$$

Applying multipoint corrections

A combination of additive and multiplicative corrections can provide for additional flexibility in minimizing the impact of the correction away from the trust region center. In other words, both additive and multiplicative corrections can satisfy local consistency, but through the combination, global accuracy can be addressed as well. Notionally, this involves a combined additive and multiplicative correction of the form:

$$f_{hi}(\mathbf{x}) = f_{lo}(\mathbf{x})B(\mathbf{x}) + A(\mathbf{x}) \quad (28)$$

The derivation is simpler, however, using a weighted sum of $\hat{f_{hi_\alpha}}$ and $\hat{f_{hi_\beta}}$

$$\hat{f_{hi_\gamma}}(\mathbf{x}) = \gamma\hat{f_{hi_\alpha}}(\mathbf{x}) + (1-\gamma)\hat{f_{hi_\beta}}(\mathbf{x}) \quad (29)$$

which ensures that $\hat{f_{hi_\gamma}}$ retains the local consistency properties of $\hat{f_{hi_\alpha}}$ and $\hat{f_{hi_\beta}}$. The γ parameter can be selected to enforce an additional matching condition, such as matching the high-fidelity function value at a nearby point $\mathbf{x_p}$ (e.g., the previous design point in the case of a successful trust region cycle, or the previously rejected iterate in the case of an unsuccessful trust region cycle). This multipoint correction then provides some control over the global accuracy of the correction. It can be shown that the following value for γ

$$\gamma = \frac{f_{hi}(\mathbf{x_p}) - \hat{f_{hi_\beta}}(\mathbf{x_p})}{\hat{f_{hi_\alpha}}(\mathbf{x_p}) - \hat{f_{hi_\beta}}(\mathbf{x_p})} \quad (30)$$

results in $\hat{f_{hi_\gamma}}(\mathbf{x_p}) = f_{hi}(\mathbf{x_p})$. As $\hat{f_{hi_\alpha}}(\mathbf{x_p}) \to f_{hi}(\mathbf{x_p})$ (the additive correction is accurate), then $\gamma \to 1$, and as $\hat{f_{hi_\beta}}(\mathbf{x_p}) \to f_{hi}(\mathbf{x_p})$ (the multiplicative correction is accurate), then $\gamma \to 0$.

It is also possible to match higher order information at $\mathbf{x_p}$ by using a first- or second-order function $\gamma(x)$. However, this results in up to a fourth-order correction in $\hat{f_{hi_\gamma}}(\mathbf{x})$, which would likely introduce more difficulties due to multimodality than it would help due to increased accuracy.

Verifying local consistency

Zeroth, first, and second-order consistency at $\mathbf{x} = \mathbf{x_c}$ ($\hat{f_{hi}}(\mathbf{x_c}) = f_{hi}(\mathbf{x_c})$, $\nabla\hat{f_{hi}}(\mathbf{x_c}) = \nabla f_{hi}(\mathbf{x_c})$, and $\nabla^2\hat{f_{hi}}(\mathbf{x_c}) = \nabla^2 f_{hi}(\mathbf{x_c})$, respectively) can be demonstrated for additive corrections by substituting Eqs. 10-12 into Eqs. 8, 24, and 26, substituting the resulting expressions into Eqs. 18-20, and then evaluating at $\mathbf{x} = \mathbf{x_c}$. Similarly, consistency for multiplicative corrections can be demonstrated by substituting Eqs. 13-15 into Eqs. 9, 25, and 27, substituting the resulting expressions into Eqs. 21-23, and then evaluating at $\mathbf{x} = \mathbf{x_c}$.

Simplifications for lower order corrections

These derivations have shown the general case of full second-order corrections. For first-order corrections, one can simply take $\nabla^2 A(\mathbf{x_c})$ and $\nabla^2 B(\mathbf{x_c})$ to be zero, in which case only zeroth and first-order consistency holds at $\mathbf{x} = \mathbf{x_c}$ ($\hat{f_{hi}}(\mathbf{x_c}) = f_{hi}(\mathbf{x_c})$ and $\nabla\hat{f_{hi}}(\mathbf{x_c}) = \nabla f_{hi}(\mathbf{x_c})$, respectively). For zeroth-order corrections, one can additionally take $\nabla A(\mathbf{x_c})$ and $\nabla B(\mathbf{x_c})$ to be zero, in which case only zeroth-order consistency holds at $\mathbf{x} = \mathbf{x_c}$ ($\hat{f_{hi}}(\mathbf{x_c}) = f_{hi}(\mathbf{x_c})$).

Approximating derivatives

To use a second-order correction when second-order information ($\nabla^2 f_{hi}(\mathbf{x_c})$, $\nabla^2 f_{lo}(\mathbf{x_c})$, or both) is not directly available, one can estimate the missing information using finite differences or approximate it through use of quasi-Newton or Gauss-Newton approximations. These procedures will often be needed to make second-order corrections practical for engineering applications.

In the finite difference case, numerical Hessians are commonly computed using either first-order forward differences of gradients using

$$\nabla^2 f(\mathbf{x}) \cong \frac{\nabla f(\mathbf{x} + h\mathbf{e}_i) - \nabla f(\mathbf{x})}{h} \quad (31)$$

to estimate the i^{th} Hessian column when gradients are analytically available, or second-order differences of function values using

$$\nabla^2 f(\mathbf{x}) \cong \frac{f(\mathbf{x}+h\mathbf{e}_i+h\mathbf{e}_j) - f(\mathbf{x}+h\mathbf{e}_i-h\mathbf{e}_j) - f(\mathbf{x}-h\mathbf{e}_i+h\mathbf{e}_j) + f(\mathbf{x}-h\mathbf{e}_i-h\mathbf{e}_j)}{4h^2} \quad (32)$$

to estimate the ij^{th} Hessian term when gradients are not directly available. This approach has the advantage of locally-accurate Hessians for each point of interest (which can lead to quadratic convergence rates in discrete Newton methods), but has the disadvantage that numerically estimating each of the matrix terms can be expensive.

Quasi-Newton approximations, on the other hand, do not reevaluate all of the second-order information for every point of interest. Rather, they accumulate approximate curvature information over time using secant updates. Since they utilize the existing gradient evaluations, they do not require any additional function evaluations for evaluating the Hessian terms. The quasi-Newton approximations of interest include the Broyden-Fletcher-Goldfarb-Shanno (BFGS) update

$$\mathbf{B}_{k+1} = \mathbf{B}_k - \frac{\mathbf{B}_k \mathbf{s}_k \mathbf{s}_k^T \mathbf{B}_k}{\mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k} + \frac{\mathbf{y}_k \mathbf{y}_k^T}{\mathbf{y}_k^T \mathbf{s}_k} \tag{33}$$

and the Symmetric Rank 1 (SR1) update

$$\mathbf{B}_{k+1} = \mathbf{B}_k + \frac{(\mathbf{y}_k - \mathbf{B}_k \mathbf{s}_k)(\mathbf{y}_k - \mathbf{B}_k \mathbf{s}_k)^T}{(\mathbf{y}_k - \mathbf{B}_k \mathbf{s}_k)^T \mathbf{s}_k} \tag{34}$$

where $\mathbf{B}_k$ is the k^{th} approximation to the Hessian $\nabla^2 f$, $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$ is the step and $\mathbf{y}_k = \nabla f_{k+1} - \nabla f_k$ is the corresponding yield in the gradients. An initial scaling of $\mathbf{y}_k^T \mathbf{y}_k / \mathbf{y}_k^T \mathbf{s}_k \mathbf{I}$ is used for $\mathbf{B}_0$ prior to the first update.[12] Both the BFGS and SR1 updates require safeguarding against numerical failures. A common safeguard for BFGS is to use the damped BFGS approach when the curvature condition $\mathbf{y}_k^T \mathbf{s}_k > 0$ is (nearly) violated.[12] However, while this is appropriate for Newton-like optimization algorithms, numerical experience indicates that the damped BFGS update significantly degrades second-order correction performance in multifidelity surrogate-based optimization, since the steps generated by the trust region cycles are not generally Newton-like and can frequently violate the curvature condition. A more effective approach has been to ignore the curvature condition and simply safeguard against small denominators in Eq. 33, skipping the update if $|\mathbf{y}_k^T \mathbf{s}_k| < 10^{-6} \mathbf{s}_k^T \mathbf{B}_k \mathbf{s}_k$. In the SR1 case, the update is similarly skipped when the denominator in Eq. 34 is small, in particular when $|(\mathbf{y}_k - \mathbf{B}_k \mathbf{s}_k)^T \mathbf{s}_k| < 10^{-6} ||\mathbf{s}_k||_2 ||\mathbf{y}_k - \mathbf{B}_k \mathbf{s}_k||_2$.

The finite difference Hessian and quasi-Newton Hessian approximation approaches are applicable to general optimization problems. In the case of a nonlinear least squares problem, the Gauss-Newton approximation

$$\nabla^2 f(\mathbf{x}) \cong \mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) \tag{35}$$

provides a third possibility where $\mathbf{J} = \nabla \mathbf{R}^T$ is the Jacobian of the residual vector $\mathbf{R}$. In this case, the Hessian approximation is instantaneous (i.e., does not require accumulation of information over time) and will improve as the residuals are driven toward zero.

Similar approaches may be used when first-order information ($\nabla f_{hi}(\mathbf{x_c})$, $\nabla f_{lo}(\mathbf{x_c})$, or both) is not directly available. To estimate gradients with finite differences, standard forward difference

$$\nabla f(\mathbf{x}) \cong \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})}{h} \tag{36}$$

or central difference approaches

$$\nabla f(\mathbf{x}) \cong \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x} - h\mathbf{e}_i)}{2h} \tag{37}$$

are employed to estimate the i^{th} component of the gradient vector. It is also possible to approximate gradients over time using a secant approximation such as Broyden's update[13]

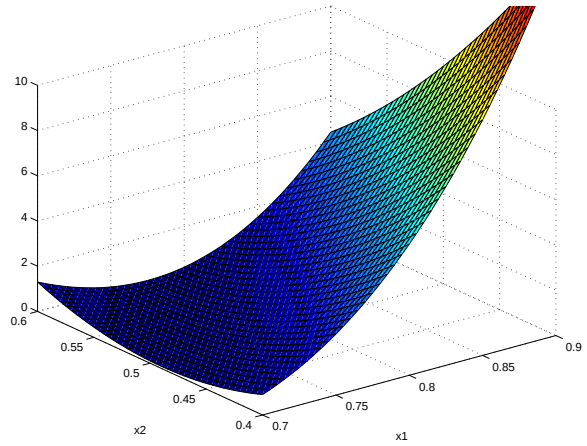
$$\mathbf{A}_{k+1} = \mathbf{A}_k + \frac{(\mathbf{y}_k - \mathbf{A}_k \mathbf{s}_k)\mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{s}_k} \tag{38}$$

where $\mathbf{A}_k$ is the k^{th} approximation to the Jacobian $\nabla \mathbf{f}^T$, $\mathbf{s}_k$ is the step as before and $\mathbf{y}_k = \mathbf{f}_{k+1} - \mathbf{f}_k$ is now the corresponding yield in the values of one or more functions. However, while this secant update approach can be effective in the solution of nonlinear equations, it is not normally recommended for optimization since local gradient accuracy is more critical in this case.

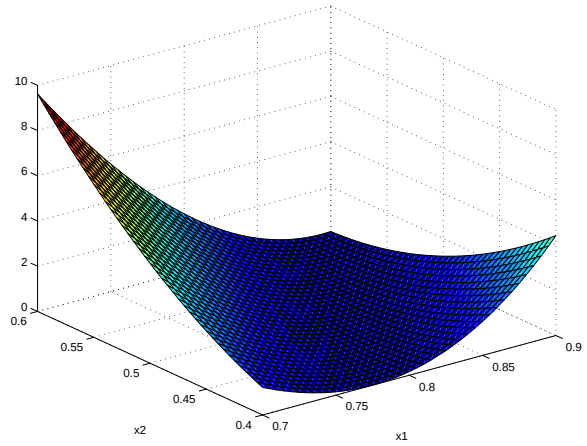
Developing physical intuition

Additive corrections add a scalar (zeroth-order), a linear function (first-order), or a quadratic function (second-order) to the low-fidelity function. This equates to a uniform translation in the zeroth-order case, a translation and rotation in the first-order case, and a translation, rotation, and uniform "bending" or "rounding" (inclusion of constant curvature) in the second-order case. Multiplicative corrections, conversely, multiply the low-fidelity function by a scalar (zeroth-order), a linear function (first-order), or a quadratic function (second-order). These corrections typically induce more skewing of the low-fidelity contours.

For example, Figure 1 shows a region of the high and low-fidelity functions taken from Example 1.1 in the Computational Experiments section (Eqs. 39-40). Correcting the low-fidelity function at the center of the region using zeroth, first, and second-order additive corrections results in Figure 2, whereas correcting the low-fidelity function using multiplicative corrections results in Figure 3. It is evident for this example that the first and second-order additive corrections result in good agreement with the high-fidelity model over the full region, whereas the accuracy of the multiplicative corrections degrades rapidly away from the center of the region. Clearly, the additional skewing of a multiplicative correction is undesirable in this example where the form of the low and high-fidelity models are relatively similar (i.e., the orders of the polynomial forms are identical, but the coefficients differ).



(a) High-fidelity function

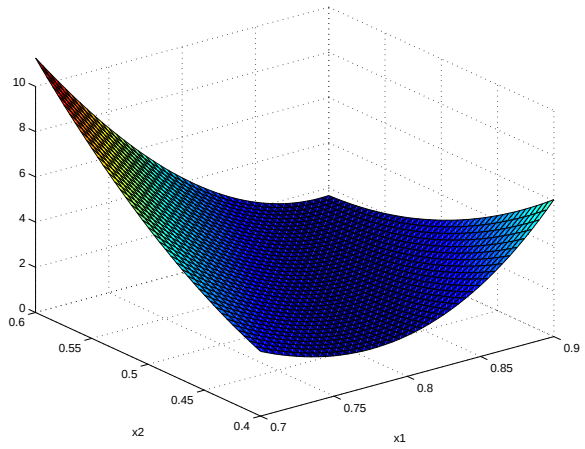


(b) Low-fidelity function

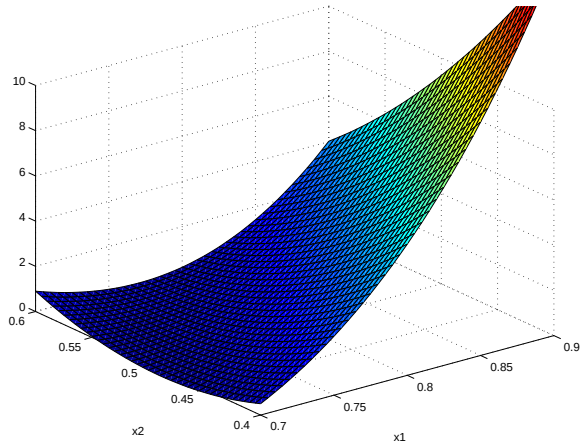
Fig. 1 Correction example 1: high and low-fidelity functions.

In cases where the form of the low and high-fidelity models are sufficiently different, then the additional skewing provided by multiplicative corrections may be desirable. For example, if the order of the high-fidelity model is greater than the low-fidelity model as in Example 2 from the Computational Experiments section (Eqs. 43-44), then one might expect better performance from a multiplicative correction. Figure 4 shows a region of these high and low-fidelity functions, Figure 5 shows the low-fidelity function with additive corrections applied, and Figure 6 shows the low-fidelity function with multiplicative corrections applied. The relative performance of the multiplicative corrections is indeed improved, and the zeroth and first-order multiplicative corrections shown in Figure 6(a,b) are noticeably more accurate than the zeroth and first-order additive corrections shown in Figure 5(a,b).

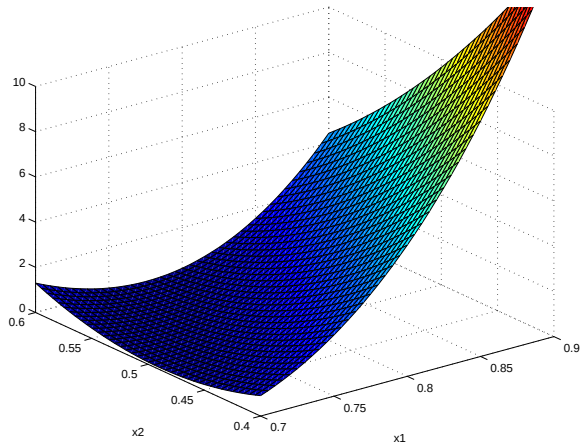
Thus, if considerable information is known about the form of the low and high-fidelity models, then it is possible to select a correction that is optimized for a particular problem. Without detailed information however, the additive corrections appear to be less susceptible to large scale inaccuracies and more consistent



(a) zeroth-order

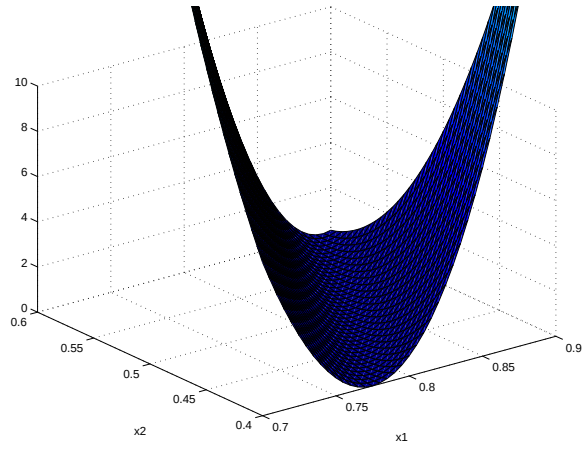


(b) first-order

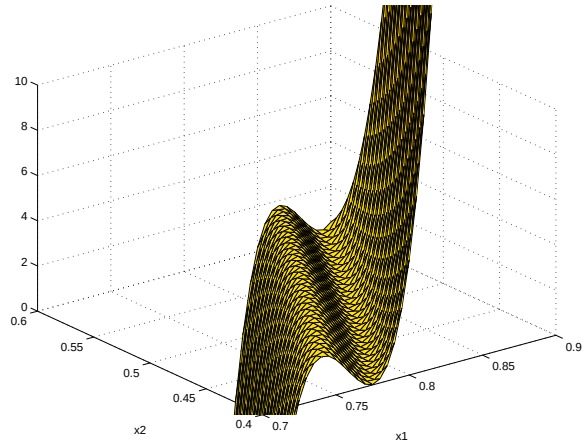


(c) second-order

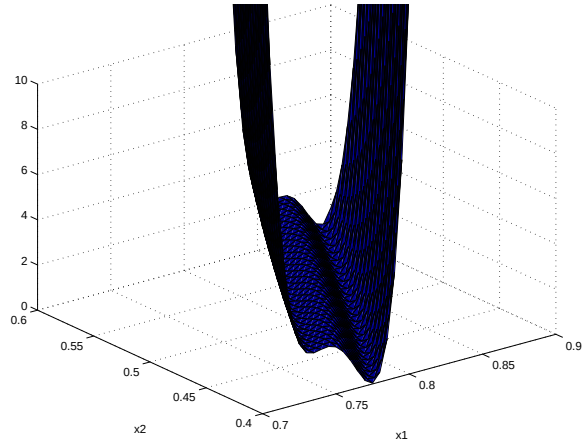
Fig. 2 Correction example 1: low-fidelity with additive corrections.



(a) zeroth-order

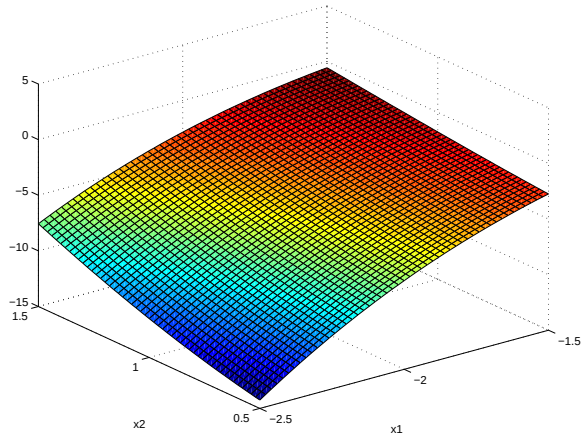


(b) first-order

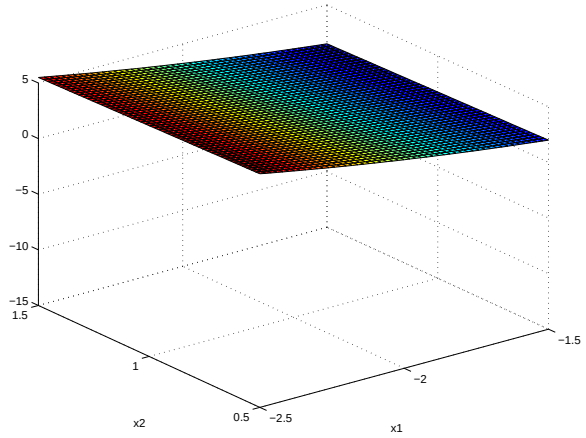


(c) second-order

Fig. 3 Correction example 1: low-fidelity with multiplicative corrections.



(a) High-fidelity function



(b) Low-fidelity function

Fig. 4 Correction example 2: high and low-fidelity functions.

in their quality.

Computational Experiments

The following computational experiments are performed using the trust-region surrogate-based optimization implementation[5] in the DAKOTA open-source software toolkit.[14] The intent is to compare the performance of additive and multiplicative corrections of different orders for both simple analytic functions as well as actual engineering problems.

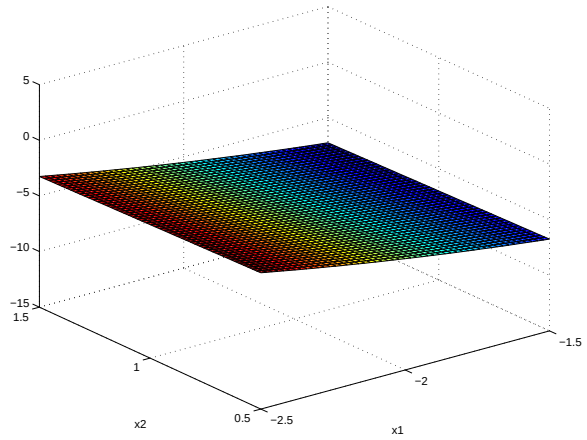
Example 1. Rosenbrock's function

Experimentation with Rosenbrock's function provided the initial motivation for the development of higher-order corrections, as first-order corrections are insufficient to achieve acceptable convergence characteristics (Rosenbrock's function is particularly difficult to solve with first-order optimization methods).

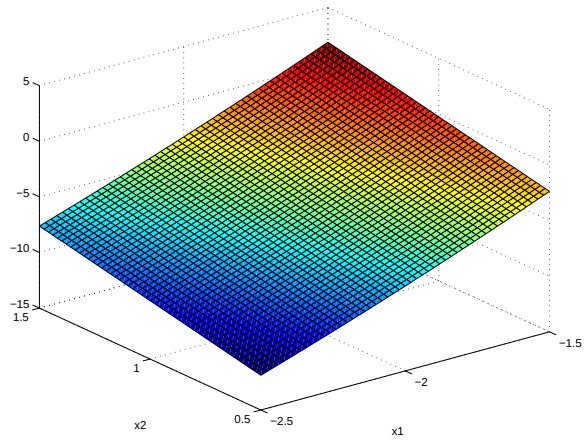
Using the standard Rosenbrock function as the "high-fidelity" function

$$f_{hi}(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2 \qquad (39)$$
$$-2 \leq \mathbf{x} \leq 2$$

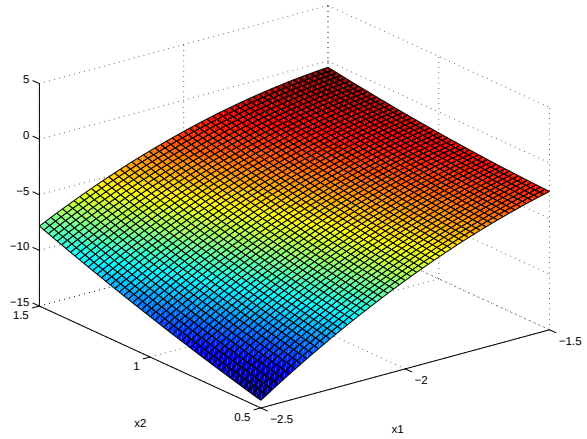
with a minimum at $\mathbf{x} = (1, 1)$, one can define several



(a) zeroth-order

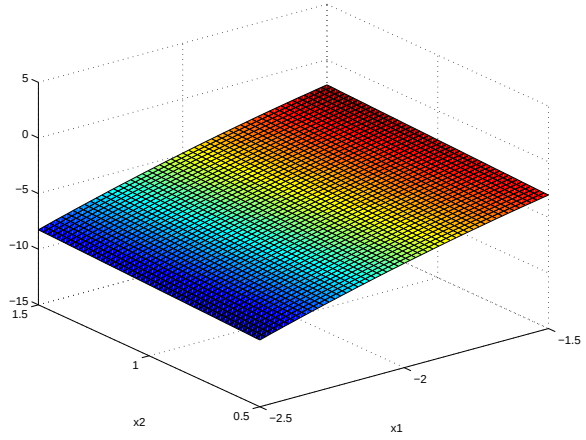


(b) first-order

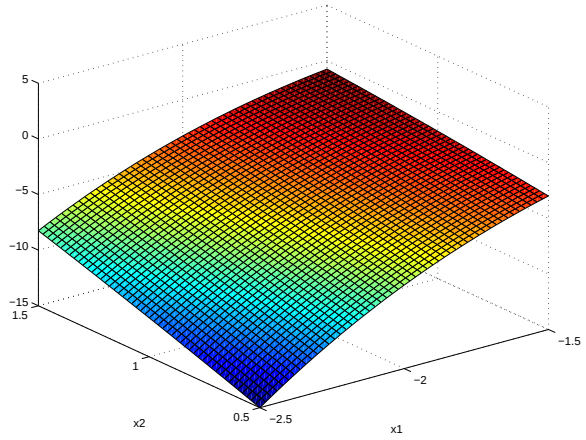


(c) second-order

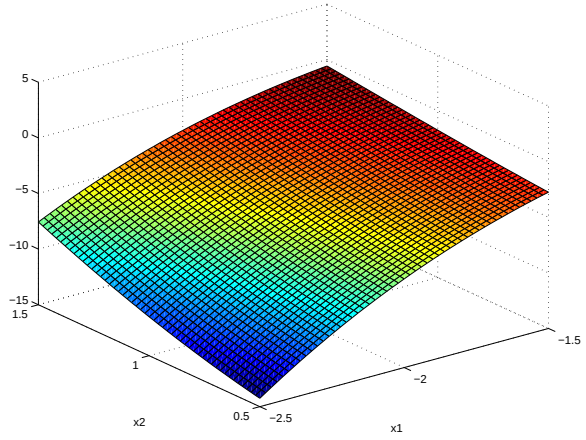
Fig. 5 Correction example 2: low-fidelity with additive corrections.



(a) zeroth-order



(b) first-order



(c) second-order

Fig. 6 Correction example 2: low-fidelity with multiplicative corrections.

"low-fidelity" Rosenbrock functions for use as inexpensive correction approach testers.

Example 1.1 Low-fidelity with offsets

This low-fidelity Rosenbrock function has ±0.2 offsets from the true function within the squared terms:

$$f_{lo}(\mathbf{x}) = 100(x_2 - x_1^2 + .2)^2 + (0.8 - x_1)^2 \qquad (40)$$

with a low-fidelity minimum at $\mathbf{x} = (0.8, 0.44)$. This modification is sufficient for the value, gradient, and Hessian of the low-fidelity function to differ from the high-fidelity values.

Table 1 displays the SBO results for each of the corrections applied to the low-fidelity Rosenbrock function starting from the point $\mathbf{x} = (-1.2, 1.0)$ with an initial trust region size of 10% of the global bounds. Duplicate function evaluations are excluded from the totals. The second-order approaches achieve hard convergence and their trust regions grow for several consecutive cycles prior to convergence. The first-order approaches terminate through a soft convergence criterion, specifically when their minimum trust region bounds are reached after a series of trust region reductions. The zeroth-order additive approach converges to the optimum of the low-fidelity model $\mathbf{x} = (0.8, 0.44)$, as would be expected with only a scalar offset, whereas the zeroth-order multiplicative approach stalls out (minimum trust region bounds are reached) short of the optimum of the low-fidelity model. This stalling behavior can occur since trust region steps taken in the direction of the optimum of the low-fidelity model (due to the poor correction) must also satisfy simple decrease in the high-fidelity model (due to the SBO step verifications). The combined/multipoint corrections do not appear to improve upon the best performing local correction, but do appear to do an acceptable job of selecting from among the local corrections (mirrors additive for zeroth-order and all second-order), although this is not always the case (mirrors multiplicative for first-order).

Figure 7 displays the convergence rates for the zeroth, first, and second-order additive and multiplicative correction approaches. The combined correction cases are not plotted, since they mostly overlap other existing curves. In Figure 7(a), quadratic convergence rates are evident in the second-order correction approaches, and additive corrections significantly outperform their multiplicative counterparts for all correction orders. Figures 7(b) and 7(c) compare the convergence rates for the finite difference, BFGS quasi, and SR1 quasi second-order corrections to the first-order and full second-order corrections for additive and multiplicative cases. The finite difference second-order convergence rate closely follows that of full second-order. The quasi-2nd-order convergence curves are delayed, exhibiting performance similar to first-order approaches early on, until sufficient curvature information is accumulated to attain improved convergence rates. Relative performance of the updates is mixed: SR1 updating outperforms BFGS updating in one case, and BFGS outperforms SR1 in the other.

Table 1 SBO correction results, low-fidelity Rosenbrock with offsets.

Correction Approach	SBO Iters	Fn Evals (LF/HF)	HF Obj Function
0th add	27	74/12	4.04
0th mult	22	125/23	4.76
0th comb	25	70/9	4.04
1st add	98	1124/195	9.29e-10
1st mult	7398	7448/12955	1.31e-05
1st comb	6995	27500/12145	5.68e-05
Full 2nd add	5	68/11	1.24e-15
Full 2nd mult	31	255/59	8.96e-15
Full 2nd comb	5	83/11	1.24e-15
FD 2nd add	5	85/23	1.53e-10
FD 2nd mult	34	303/124	3.18e-10
FD 2nd comb	5	75/23	8.27e-11
BFGS 2nd add	41	261/75	2.53e-07
BFGS 2nd mult	43	310/84	1.13e-13
BFGS 2nd comb	41	242/73	3.52e-07
SR1 2nd add	23	158/42	8.29e-15
SR1 2nd mult	93	389/151	1.61e-09
SR1 2nd comb	23	148/42	4.73e-15

The initial trust region size is fixed for all methods to provide a consistent comparison. However, with a larger initial trust region of 50%, the multifidelity SBO approach with second-order additive correction converges in as few as three high-fidelity function value/gradient/Hessian evaluations. Compared to a single-fidelity full-Newton optimization of Rosenbrock's function, the number of high-fidelity evaluations is reduced by approximately an order of magnitude. Therefore, if the low-fidelity evaluations were significantly less expensive than the high-fidelity evaluations, the multifidelity SBO approach would have significant savings over the single-fidelity approach.

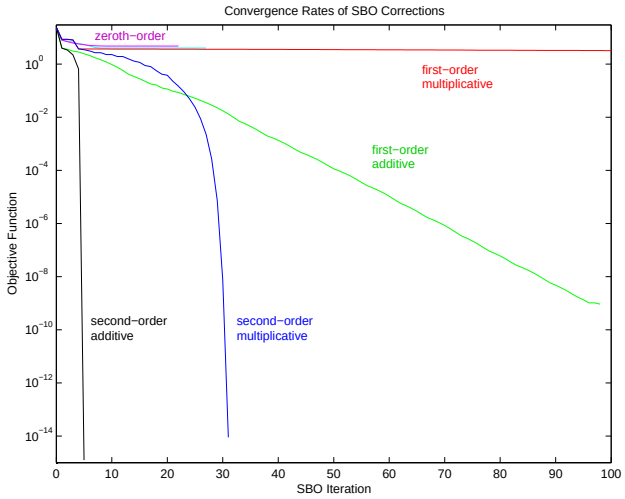
Example 1.2. Low-fidelity with scalings

This low-fidelity Rosenbrock function has 1.25 scalings from the true function within the squared terms:

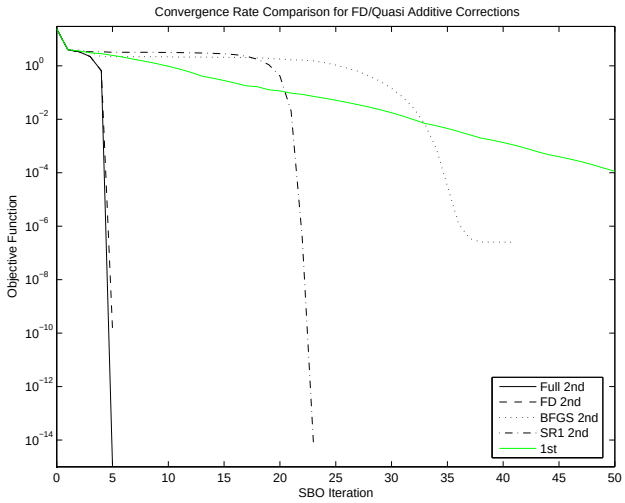
$$f_{lo}(\mathbf{x}) = 100(1.25x_2 - x_1^2)^2 + (1 - 1.25x_1)^2 \qquad (41)$$

with a low-fidelity minimum at $\mathbf{x} = (0.8, 0.512)$. The purpose of this example is to explore a modification which is more multiplicative in nature (for which multiplicative corrections may be more competitive).

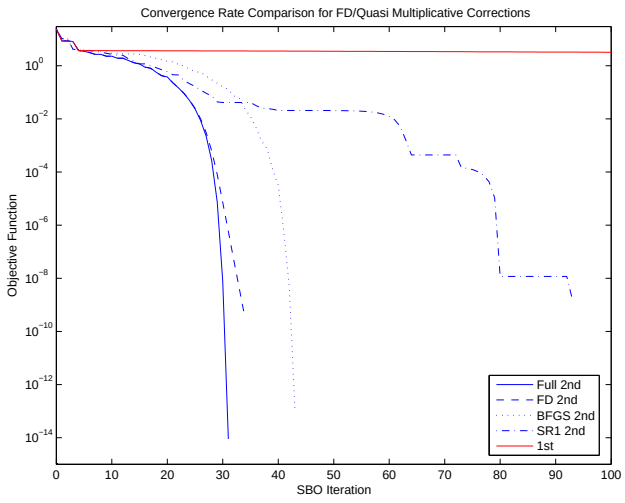
Table 2 displays SBO results for the same cases investigated in Example 1.1. The second-order approaches again achieve hard convergence, the first-order approaches again terminate through the minimum trust region soft convergence criterion, and the zeroth-order approaches stall out (minimum trust region bounds are reached) short of the optimum of



(a) Zeroth, first, and full second-order corrections



(b) Additive corrections: comparison of FD/quasi



(c) Multiplicative corrections: comparison of FD/quasi

Fig. 7 Convergence comparison for zeroth, first, and second-order corrections to the low-fidelity Rosenbrock function with offsets.

Table 2 SBO correction results, low-fidelity Rosenbrock with scalings.

Correction Approach	SBO Iters	Fn Evals (LF/HF)	HF Obj Function
0th add	25	134/26	0.468
0th mult	26	161/27	0.461
0th comb	25	153/26	0.468
1st add	861	11347/1620	7.58e-08
1st mult	7252	7404/12713	1.22e-05
1st comb	5439	14564/9505	3.62e-03
Full 2nd add	17	183/34	1.82e-09
Full 2nd mult	42	331/76	2.59e-12
Full 2nd comb	56	603/96	3.65e-09
FD 2nd add	17	187/68	4.58e-09
FD 2nd mult	45	369/153	7.59e-10
FD 2nd comb	30	272/99	2.67e-09
BFGS 2nd add	74	1472/133	5.47e-10
BFGS 2nd mult	87	396/154	1.38e-13
BFGS 2nd comb	292	1381/514	1.68e-14
SR1 2nd add	207	1560/343	2.27e-07
SR1 2nd mult	102	467/176	4.25e-11
SR1 2nd comb	224	1059/375	7.26e-05

the low-fidelity model (due to the SBO verifications for decrease in the high-fidelity model). The combined/multipoint corrections are not as effective in this problem as for Example 1.1, and are outperformed by both local corrections in the majority of the cases.

Figure 8 displays the convergence rates for the zeroth, first, and second-order additive and multiplicative corrections. In Figure 8(a), superior convergence rates are again evident in the second-order correction approaches. While the multiplicative corrections are slightly more competitive relative to Example 1.1, they are still outperformed by the additive corrections. Figures 8(b) and 8(c) show similar performance as in Example 1.1. The finite difference second-order convergence rate again mirrors the full second-order rate, and the quasi second-order corrections outperform the first-order but display a lag relative to second-order. For this problem, BFGS updating outperforms SR1 updating.

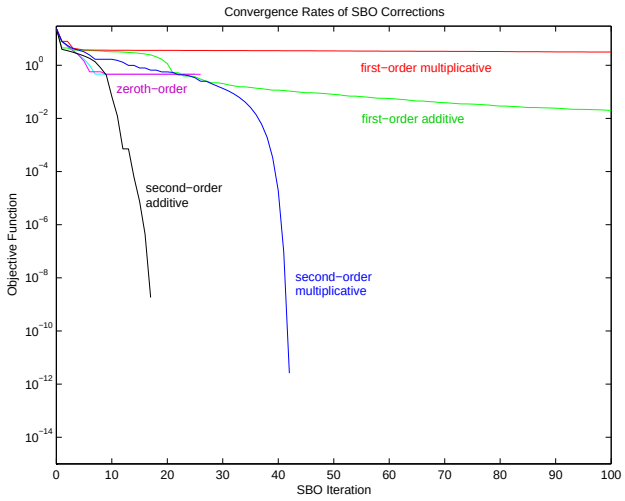
Example 1.3. Constant low-fidelity

This low-fidelity Rosenbrock function is a constant:

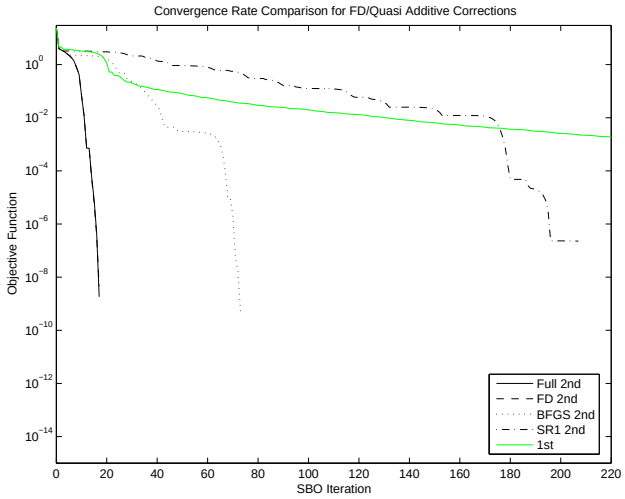
$$f_{lo}(\mathbf{x}) = 100 \tag{42}$$

with no unique minimum. The purpose of this example is to provide insight into the relative importance of capturing representative features of the high-fidelity problem within the low-fidelity model by examining a poor low-fidelity model.

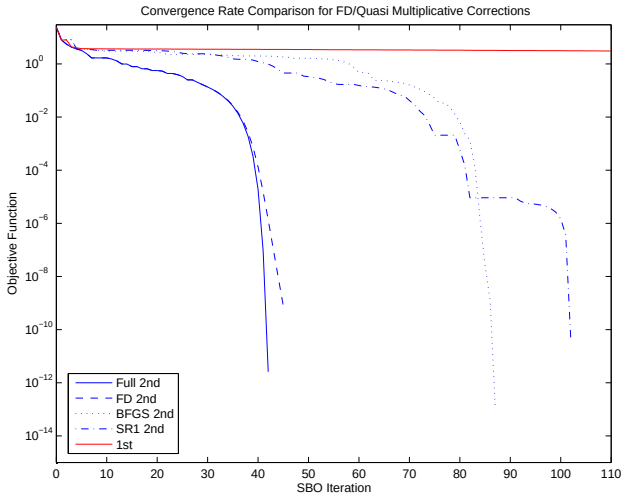
It can be shown that, for $\nabla f_{lo}(\mathbf{x}) = \nabla^2 f_{lo}(\mathbf{x}) = 0$ and $f_{lo}(\mathbf{x}) = f_{lo}(\mathbf{x_c})$, the additive corrections in Eqs. 18-20 and the multiplicative corrections in Eqs. 21-23 simplify to identical expressions, those corresponding to a second-order Taylor series of $f_{hi}(\mathbf{x})$



(a) Zeroth, first, and full second-order corrections



(b) Additive corrections: comparison of FD/quasi



(c) Multiplicative corrections: comparison of FD/quasi

Fig. 8 Convergence comparison for zeroth, first, and second-order corrections to the low-fidelity Rosenbrock function with scalings.

Table 3 SBO correction results, constant low-fidelity Rosenbrock.

Correction Approach	SBO Iters	Fn Evals (LF/HF)	HF Obj Function
0th	17	2/1	24.2
1st	7363	7388/12891	1.28e-05
Full 2nd	24	135/45	3.32e-12
FD 2nd	26	171/95	2.31e-10
BFGS 2nd	59	284/104	2.06e-11
SR1 2nd	121	383/211	1.12e-10

centered at $\mathbf{x_c}$. Thus, the additive and multiplicative correction results are identical for this example and correspond to the performance of a single-fidelity SBO based solely on high-fidelity data. The combined correction is not well-defined in this case (Eq. 30), so numerical safeguarding selects γ to be unity (an additive correction).

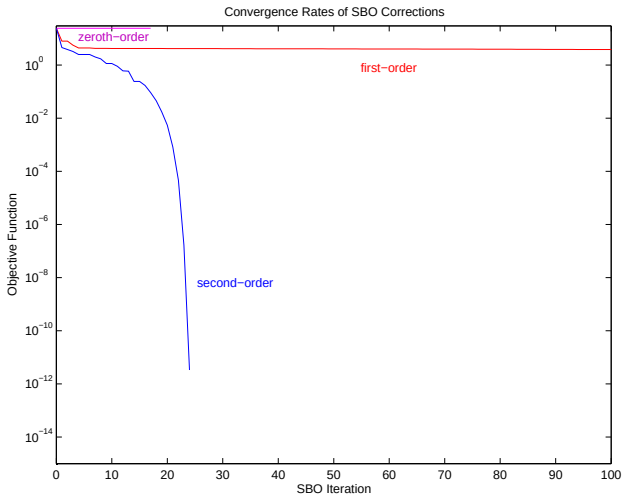
Table 3 displays the SBO results for each of the correction approaches with additive, multiplicative, and combined corrections collapsed to a single entry. Second-order correction again achieves hard convergence, first-order correction again terminates through the minimum trust region soft convergence criterion, and zeroth-order correction fails to make any progress since each approximate optimization cycle terminates immediately with zero gradients (which terminates the zeroth-order SBO process through a soft convergence criterion after a sequence of consecutive failures).

Figure 9 displays the convergence rates for the zeroth, first, and second-order correction approaches. In Figure 9(a), a superior convergence rate is again evident for second-order correction. Figure 9(b) compares the convergence rates for the finite difference and quasi-second-order correction variants to the first-order and full second-order corrections. The finite difference second-order convergence rate again mirrors the full second-order rate, and the quasi second-order corrections again outperform the first-order while displaying a lag relative to second-order. BFGS updating again outperforms SR1 updating.

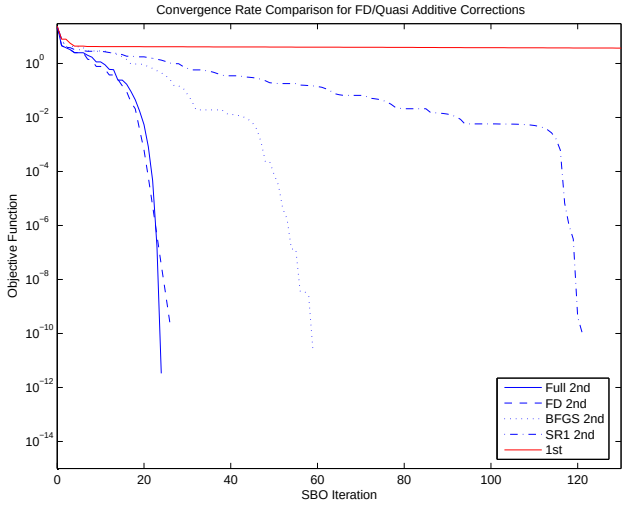
When comparing Figure 7(a) and Figure 8(a) from the first two examples with Figure 9(a) in this example, it is evident that the additive corrections do benefit from the capture of high-fidelity features in the low-fidelity model. The trend is less conclusive with the multiplicative corrections. This provides a measure of validation for the use of multi-fidelity surrogate-based optimization approaches as opposed to using single-fidelity surrogate-based optimization approaches based solely on approximations of the high-fidelity model.

Example 2. Polynomial product

The purpose of this example is to explore whether multiplicative corrections may outperform additive corrections when the high-fidelity model (Eq. 44) is



(a) Zeroth, first, and full second-order corrections



(b) Comparison of FD/quasi corrections

Fig. 9 Convergence rates for zeroth, first, and second-order corrections to the constant low-fidelity Rosenbrock function.

of higher order than the low-fidelity model (Eq. 43).

$$f_{lo}(\mathbf{x}) = x_1^2 - \frac{x_2}{2} \quad (43)$$

$$f_{hi}(\mathbf{x}) = (x_1 + \frac{x_2^2}{2}) f_{lo}(\mathbf{x}) \quad (44)$$
$$-5 \leq \mathbf{x} \leq 5$$

The high-fidelity global minimum occurs at $\mathbf{x} = (-5., -.0997)$ and the low-fidelity minimum occurs at $\mathbf{x} = (0., 5.)$.

Table 4 displays the SBO results for each of the corrections applied to the low-fidelity function starting from the point $\mathbf{x} = (-2., 1.)$ with an initial trust region size of 10% of the global bounds (Figures 4-6 display the high-fidelity, low-fidelity, and corrected low-fidelity functions for this initial trust region). For this case, specifically tailored for multiplicative corrections, the multiplicative corrections do generally outperform the additive corrections, although not by a large margin. The second-order approaches are successful and terminate through hard convergence, the first-order approaches are successful and terminate through soft convergence criteria, and the zeroth-order approaches are unsuccessful. The combined/multipoint corrections mirror the effective multiplicative corrections, and even outperform both the additive and multiplicative corrections in the first-order case. The initial trust region size is set consistently for each of the methods, but is too restrictive for the second-order correction methods. With a larger initial trust region, additive second-order will converge in as few as 4 iterations and multiplicative second-order will converge in only 1 iteration (exact for this problem).

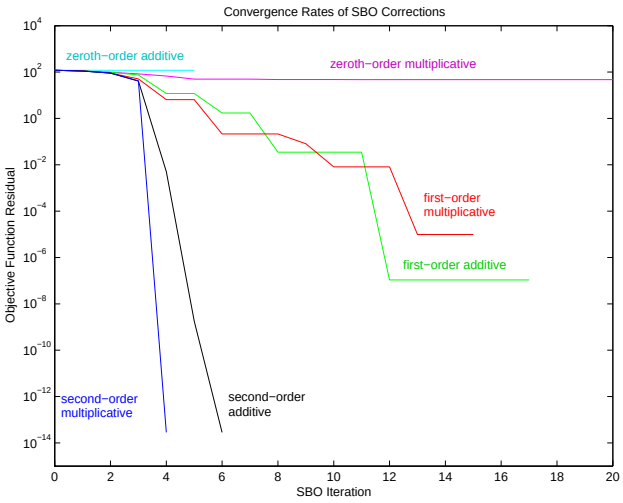
Table 4 SBO correction results, polynomial product test problem.

Correction Approach	SBO Iters	Fn Evals (LF/HF)	HF Obj Fn Residual
0th add	5	7/6	120.
0th mult	20	21/20	47.1
0th comb	5	7/6	120.
1st add	17	18/24	1.07e-07
1st mult	15	15/21	9.83e-06
1st comb	10	15/19	1.78e-12
Full 2nd add	6	21/13	2.84e-14
Full 2nd mult	4	18/9	2.84e-14
Full 2nd comb	4	18/9	2.84e-14
FD 2nd add	6	29/27	2.84e-14
FD 2nd mult	5	28/23	2.84e-14
FD 2nd comb	5	28/23	2.84e-14
BFGS 2nd add	9	20/19	7.18e-12
BFGS 2nd mult	9	25/19	1.59e-12
BFGS 2nd comb	9	24/19	2.84e-14
SR1 2nd add	9	22/19	2.84e-14
SR1 2nd mult	8	25/17	7.11e-14
SR1 2nd comb	8	22/17	2.84e-14

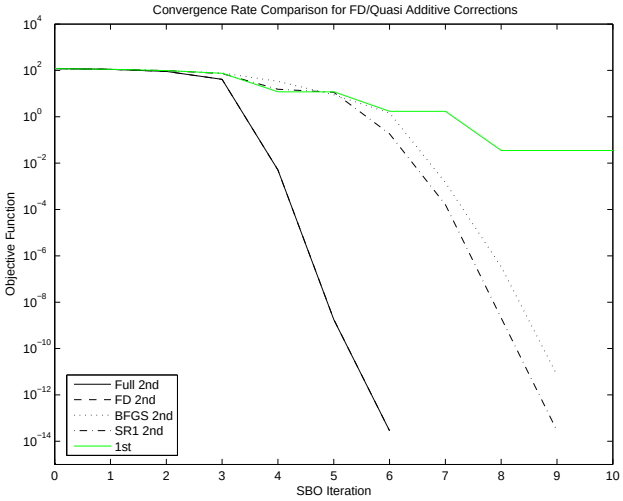
Figure 10 displays the convergence rates for the zeroth, first, and second-order additive and multiplicative correction approaches. In Figure 10(a), superior convergence rates are again evident in the second-order correction approaches. Figures 10(b) and 10(c) compare the convergence rates for the finite difference, BFGS quasi, and SR1 quasi second-order corrections to the first-order and full second-order corrections. The finite difference second-order convergence rates again mirror the full second-order rates (their iteration histories in Figure 10(b) are identical), and the quasi second-order corrections outperform the first-order but display a lag relative to second-order. For this problem, SR1 updating slightly outperforms BFGS updating.

Example 3. Burgers Boundary Control

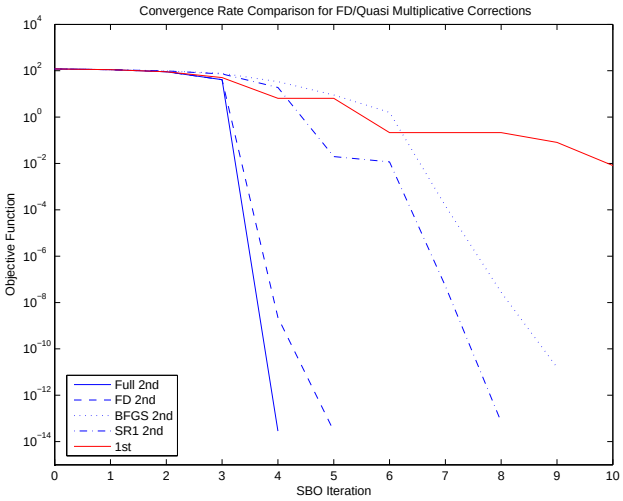
In this example, the SBO approach is applied to a model, nonlinear PDE constrained optimization prob-



(a) Zeroth, first, and full second-order corrections



(b) Additive corrections: comparison of FD/quasi



(c) Multiplicative corrections: comparison of FD/quasi

Fig. 10 Convergence comparison for zeroth, first, and second-order corrections to the polynomial product test problem.

lem. The PDE is the transient, viscous Burgers equation

$$y_{,t} + (y^2/2)_{,x} - (\nu y_{,x})_{,x} = 0 \tag{45a}$$

for $(t,x) \in (0,1] \times [0,1]$ subject to the boundary conditions

$$\nu y_{,x} + y = 0 \quad \text{at } x = 0 \tag{45b}$$

$$\nu y_{,x} - y = u \quad \text{at } x = 1 \tag{45c}$$

and with initial condition

$$y(0,x) = \sin(\pi \tan(c_s(2x-1))/\tan(c_s))\,. \tag{45d}$$

For the examples shown here, $\nu = 0.01$, $c_s = 1.3$, and u is the design (control) variable at the right boundary. The objective functional is

$$J(u,y) = \alpha\frac{u^2}{2} + \int_0^1 \int_0^1 \frac{1}{2}\left(y(t,x) - \widehat{y}(x)\right)\,dx\,dt \tag{46}$$

where the penalty on the magnitude of the control is $\alpha = 1/1000$ and the target solution $\widehat{y}(x) = y(0,x)$. The optimization problem is then

$$\min_{u \in \mathcal{U}} J(u,y) \tag{47}$$

such that $y(t,x)$ satisfies (Eq. 45) and $\mathcal{U} = I\!R$. Obviously, this model problem is far simpler than the large-scale target applications for which an SBO approach is appropriate. However, the small size of this one-dimensional PDE allows us to explore the performance of the SBO approaches while still retaining some of the salient features — nonlinearity and transients — typical of many large-scale applications.

Equation (45a) is discretized using a discontinuous Galerkin (DG) method in space with fourth-order explicit Runge-Kutta in time.[15] In addition to providing a robust, stable, and high-order accurate method for solving conservation laws, DG also readily allows for multi-fidelity modeling where low-order models can be constructed for the same mesh topology by reducing the local polynomial order on each element. In this example, our high-fidelity model is based on a 40 element discretization using quadratic polynomial ($p = 2$) representations on each element while the low-order models use the same mesh, but with $p = 1$ or $p = 0$ polynomials on each element and with larger time-steps commensurate with the stability limits of RK-4. Figure 11 shows the variation, with u, of the objective function value, gradient, and Hessian for each model. The objective function value for the $p = 1$ model is nearly indistinguishable from that of the $p = 2$ "truth" model although there are observable differences between the two models for the gradient and Hessian. However, the lower-fidelity model constructed of a $p = 0$ (i.e. piecewise constant) solution representation shows differences even in the objective

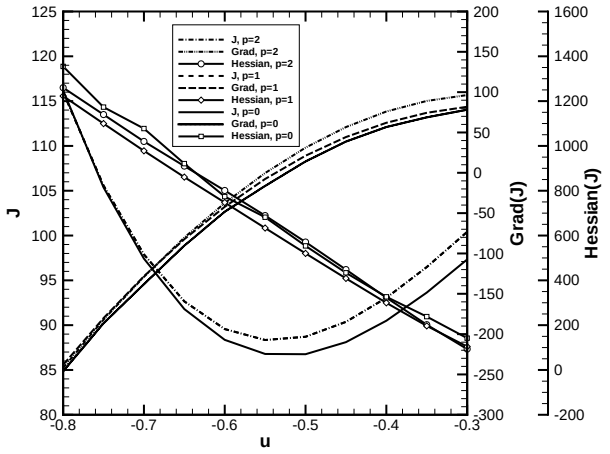


Fig. 11 Comparison of objective function, gradient, and Hessian values for Burgers equation multi-fidelity models

function value, including the value and location of the minimum.*

From the structure of the optimality landscape in Fig. 11, we have applied the additive SBO correction approach using 0^{th}, 1^{st}, and 2^{nd} order corrections. All gradient and Hessian evaluations were performed using finite difference approximations (first-order forward for the gradient and second-order central for the Hessian).

SBO with $p = 0$ model

With the $p = 0$ model, Fig. 11 shows that the minimum is clearly shifted with respect to the high-fidelity model $p = 2$. This indicates that the 0^{th}-order SBO model will not be effective and this is confirmed by the SBO results. Zeroth order correction with the $p = 0$ model was only able to perform one useful update (i.e. it found the optimal solution to the $p = 0$ model) but was then unable to further improve the solution of the high-fidelity model. Using 1^{st}-order additive SBO corrections allowed for three successful SBO iterations with a final high-fidelity gradient of $O(10^{-4})$. However, further SBO iterations with the 1^{st}-order correction make no further improvement in the solution of the high-fidelity problem. Finally, with 2^{nd}-order corrections, the gradient of the high-fidelity model converges to $O(10^{-6})$, our convergence tolerance, in four iterations.

SBO with $p = 1$ model

We now consider the $p = 1$ model which provides a more accurate prediction of the objective function, although there are still noticeable differences in the gradient and Hessian (see Fig. 11). For this model with 0^{th}-order additive corrections, two useful iterations were performed with a slightly better final solution than using the $p = 0$ model. The performance of the 1^{st}-order additive correction for the $p = 1$ model was similar to that achieved with the $p = 0$ model with the gradient error in the high-fidelity solution saturating at $O(10^{-4})$. However, with 2^{nd}-order corrections, the added fidelity of the $p = 1$ model resulted in $O(10^{-6})$ convergence in only two iterations.

Summary

We conclude the discussion of our SBO results for Burgers equation with the caveat that the optimization problem (Eq. 47) is too "easy" to see a dramatic difference in the number of function evaluations or iterations required for the different SBO corrections. This is primarily due to the fact that there is only one control variable u and the objective function is nearly quadratic (see Fig. 11). Nevertheless, this example does highlight several important issues:

- The flexibility of DG methods provides a convenient means of constructing discretization-based low-order models through p de-refinement without requiring changes to the mesh topology.

- For this problem, additive 1^{st}-order corrections within the SBO framework provide solutions within $O(10^{-4})$ for the high-fidelity problem in only 3 iterations. While further improvement in the solution was not possible, this may be sufficiently converged for many engineering applications.

- With 2^{nd}-order corrections, the SBO framework was able to compute accurate solutions within 2–4 SBO iterations (depending on the fidelity of the model). However, this comes at the expense of requiring Hessian information, which we computed using finite-differencing of the finite-difference gradient. This approach is only practical for very small control spaces and more complex problems will require either finite differencing of adjoint-based gradients or quasi-Newton approximations for second-order information.

We plan to explore these issues further and extensions of this problem to include multiple control variables (e.g. a polynomial parameterization of an unsteady control) are currently underway.

Conclusions

This paper has demonstrated that second-order corrections can lead to more desirable convergence characteristics in model-hierarchy surrogate-based optimization than the currently popular first-order correction approaches. Additive corrections are shown to be preferable to multiplicative corrections in almost all cases. For a test problem specifically constructed for

*Note the small oscillations in the $p = 0$ Hessian are due to the use of the Lax-Freidrichs flux in our DG method which is not differentiable at $y = 0$. Using a smoothed flux, such a Roe with entropy fix, would prevent such oscillations.

the benefit of multiplicative corrections, the multiplicative corrections do demonstrate a slight advantage, but not by a significant amount.

Multipoint corrections, which combine additive and multiplicative corrections in an adaptive way, are shown to be marginally effective when an additional matching condition of the previous design iterate is chosen. While the information from the previous design iterate is freely available, this is likely not the best choice as it biases the global accuracy of the correction in the direction the algorithm has already been, not where it is going. Forward-looking matching conditions, while not freely available, will likely be more effective.

Since full second-order information is not commonly available, finite difference and quasi-Newton second-order corrections have also been developed and demonstrated. The finite difference second-order corrections require additional function evaluations to estimate the second derivatives, but the convergence characteristics closely follow that of the full second-order corrections. The quasi-Newton second-order corrections do not require any additional function evaluations and consistently outperform the first-order corrections. They do however exhibit a convergence lag relative to the full and finite difference second-order corrections as they accumulate curvature information from the sequence of gradients. Both the BFGS and SR1 quasi-Newton updates were effective, with BFGS exhibiting a slight performance advantage over SR1 on average.

Overall, the introduction of new full, finite difference, and quasi-second-order correction approaches should assist in extending the performance of multifidelity surrogate-based optimization algorithms and improve their relative performance with respect to competing single-fidelity approaches.

Acknowledgments

The authors gratefully acknowledge the assistance of Natalia Alexandrov of NASA Langley and Michael Lewis of the College of William and Mary in sharing their expertise in surrogate-based optimization theory. The authors would also like to express their thanks for informative discussions with Victor Perez of the University of Notre Dame and Sandia, Shawn Gano of the University of Notre Dame, John Dennis of Rice University and Boeing, and David Gay of Sandia.

References

[1]Schmit, L. A., Jr., and Miura, H., "Approximation Concepts for Efficient Structural Synthesis," NASA CR-2552, March 1976.

[2]Kamat, M.P., ed., "Chapter 4. Function Approximations" in *Structural Optimization: Status and Promise*, Progress in Astronautics and Aeronautics, Volume 150, American Institute of Aeronautics and Astronautics, 1993.

[3]Sobieszczanski-Sobieski, J., and Haftka, R.T., "Multidisciplinary Aerospace Design Optimization: Survey of Recent Developments," *Structural Optimization*, 14(1), 1997, pp. 1-23.

[4]Alexandrov, N. M., Dennis, J. E., Jr., Lewis, R. M., and Torczon, V., "A trust region framework for managing the use of approximation models in optimization", *Structural Optimization*, 15 (1998), pp. 16-23.

[5]Giunta, A.A., and Eldred, M.S., "Implementation of a Trust Region Model Management Strategy in the DAKOTA Optimization Toolkit," paper AIAA-2000-4935 in *Proceedings of the 8th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, Long Beach, CA, September 6-8, 2000.

[6]Perez, V.M., Renaud, J. E., and Watson, L. T., "Reduced Sampling for Construction of Quadratic Response Surface Approximations Using Adaptive Experimental Design," *Proceedings of the 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, AIAA Paper 2002-1587, Denver, CO, April 22-25.

[7]Haftka, R. T., "Combining Global and Local Approximations," *AIAA Journal*, Vol. 29, No. 9, 1991, pp. 1523-1525.

[8]Lewis, R. M., and Nash, S. N., "A Multigrid Approach to the Optimization of Systems Governed by Differential Equations," AIAA Paper 2000-4890, 2000.

[9]Rodriguez, J. F., Renaud, J. E., and Watson, L. T., "Convergence of Trust Region Augmented Lagrangian Methods Using Variable Fidelity Approximation Data," *Structural Optimization*, vol. 15, pp. 1–7, 1998.

[10]Alexandrov, N. M., Lewis, R. M., Gumbert, C. R., Green, L. L., and Newman, P. A., "Optimization with Variable-Fidelity Models Applied to Wing Design," paper AIAA-2000-0841, 38th Aerospace Sciences Meeting and Exhibit, Reno, NV, 2000.

[11]Perez, V. M., Eldred, M. S., and Renaud, J. E., "Solving the Infeasible Trust-Region Problem Using Approximations," paper AIAA-2004-4312 in *Proceedings of the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Albany, NY, Aug. 30–Sept. 1, 2004.

[12]Nocedal, J., and Wright, S.J., *Numerical Optimization*, Springer, New York, 1999.

[13]Dennis, J.E., Jr., and Schnabel, R.B., *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, NJ, 1983.

[14]Eldred, M.S., Giunta, A.A., van Bloemen Waanders, B.G., Wojtkiewicz, S.F., Jr., Hart, W.E., and Alleva, M.P., "DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis. Version 3.0 Users Manual." Sandia Technical Report SAND2001-3796, April 2002.

[15]Chen, G., and Collis, S.S., "Toward Optimal Control of Aeroacoustic Flows using Discontinuous Galerkin Discretizations," paper AIAA-2004-0364 in *Proceedings of the 2004 AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, January 5–8, 2004.