Multi-Fidelity, Gradient-enhanced, and Locally Optimized Sparse Polynomial Chaos and Kriging Surrogate Models Applied to Benchmark Problems

Markus P. Rumpfkeil *
University of Dayton, Ohio, 45469, USA

Philip S. Beran †
Air Force Research Laboratory, Wright-Patterson Air Force Base, Ohio, 45433, USA

In this paper, multi-fidelity, gradient-enhanced, and locally optimized sparse polynomial chaos expansion (PCE) and kriging surrogate models are constructed in lieu of solely using computationally expensive high-fidelity engineering analyses. Once an accurate surrogate model is built, it can be used for evaluating a large number of designs for design space exploration or for uncertainty quantification. To demonstrate that accurate multi-fidelity, gradient-enhanced, and locally optimized surrogate models can be obtained at lower computational cost than basic high-fidelity ones, a number of benchmark problems are employed. These include polynomial and non-polynomial analytical functions of arbitrary dimension, a heterogeneous non-polynomial analytical function in one, two, and three dimensions as well as a coupled spring-mass-system leading to a linear system of ODEs of arbitrary dimension.

I. Motivation and Background

The Air Force Research Laboratory’s Multidisciplinary Science and Technology Center is currently investigating conceptual design processes and computing frameworks that could significantly impact the design of next-generation aircraft. These aircraft concepts will be designed to meet the ever-growing Air Force requirements for mission capability, combat survivability, and lifetime sustainability and require a multidisciplinary design and analysis approach. This approach makes it possible to capture the complex, often coupled, physical phenomena present in the operating environment such as nonlinear aeroelastic, aerodynamic, and thermal-structural effects. It also allows researchers to exploit these effects and their interactions to achieve advanced aircraft capabilities and configurations otherwise unattainable. However, these coupled analyses tend to be computationally very expensive, which poses a huge challenge since typically a large number of configurations must be analyzed. Thus, for inclusion in the overall design routine, the analysis must balance a trade-off between the fidelity of the solution and its computational time.

In order to save computational time during design space exploration and optimization or for uncertainty quantification the use of surrogate models is a very attractive option. The idea of a surrogate model is to replace expensive function evaluations with an approximate but inexpensive functional representation which can be probed exhaustively. Especially the polynomial chaos expansions (PCEs) and the kriging model have been widely used. The kriging construction process can be enhanced by a dynamic training point selection process rather than only specifying the sample size at the beginning and picking the training points through latin hypercube sampling. This is similar to the concept of expected improvement (EI) when performing optimizations with a kriging model where a potential for improvement is used which considers both estimated function values and uncertainties in the surrogate model, thereby keeping the balance between global and local search performance.

*Associate Professor, Hans von Ohain Endowed Chair, Department of Mechanical and Aerospace Engineering, Markus.Rumpfkeil@udayton.edu, Associate Fellow AIAA
†Principal Research Aerospace Engineer, Associate Fellow AIAA
Because an efficient gradient evaluation method based on adjoint formulations is available, the introduction of gradient information within surrogate models as additional training data has also attracted attention. The reason for this is that, for computational high-fidelity applications targeting a single output objective, the effort for computing the full gradient is, thanks to adjoint techniques, comparable to the effort of computing the objective function itself. Therefore, as the number of design inputs, $M$, increases, using the output function and its derivative information is appealing, because it provides $M + 1$ pieces of information for roughly the cost of two function evaluations. Thus, one can reasonably expect to have to compute the output function overall far fewer times to obtain an accurate surrogate model when using gradient information. While the computational effort for adjoint techniques is proportional to the number of output objectives this should not limit the applicability to aircraft design since the number of output objectives is usually relatively small. Gradient enhanced polynomial chaos\textsuperscript{3,35,12,13} as well as kriging\textsuperscript{14–17} models have been developed in the surrogate model community and have shown very beneficial results.

PCE\textsuperscript{13,18–21} and kriging\textsuperscript{17,22–25} models also support the usage of both high- and low-fidelity training points. The general idea is to combine trends from inexpensive, low-fidelity (LF) data (e.g., coarser meshes, less sophisticated models) with a fitting of the high-fidelity (HF) data (e.g., finer meshes, better models, experimental data). Thus, the trend of the unknown function underlying the intensively sampled LF data needs to be mapped to the usually less intensively sampled HF data, for example, via the popular correction-based method.\textsuperscript{25} The correction is called bridge function, scaling function, or calibration, and can be multiplicative,\textsuperscript{26} additive\textsuperscript{27,28} or hybrid multiplicative/additive.\textsuperscript{13,18,19,21}

Another issue may be the limited modeling flexibility of a single global model when there is heterogeneity in the underlying function. In a kriging context this means if an underlying function shows varying response behavior a stationary covariance structure will result in low quality prediction and overly conservative expected mean squared errors. This effect can be amplified by data collected adaptively and unevenly. Thus, utilizing a non-stationary kriging methodology via locally-optimized covariance holds promise as demonstrated by Clark and Bae,\textsuperscript{29,30} Liem et al.\textsuperscript{31} and Rumpfkeil et al.\textsuperscript{32,33} Lastly, the compressed sensing (CS) theory\textsuperscript{34–37} has shown great potential to reduce the curse of dimensionality for PCEs depending on the sparsity of the underlying solution or the decay rate of the PCE coefficients. CS is a developing research topic in the field of signal processing, applied mathematics, computer science and statistics and offers a framework to accurately, or even exactly, recover a sparse signal from a set of incomplete observations.\textsuperscript{37–39} The method can efficiently reconstruct a sparse signal such that the number of required measurements are much lower than the cardinality of the signal. In other words, CS aims at selecting a small number of basis polynomials with great impact on the model response.\textsuperscript{40}

In summary, the general goal of this work is to build highly accurate surrogate models at small overall computational cost assuming that obtaining high-fidelity training point information is the dominant cost factor. The strategy to achieve this is to enhance surrogate modeling techniques by adaptively selecting training points as well as utilizing derivative information and to employ lower fidelity information as well as locally optimized surrogate models and compressed sensing. Section II describes the construction process for multi-fidelity kriging models, Section III outlines the building of multi-fidelity sparse polynomial chaos expansion (MFSPCE) models and Section IV the agglomeration of locally optimized surrogates (ALOS). The global accuracy of these surrogate models is demonstrated on benchmark problems which are described in Section V and results are presented in Section VI. Section VII concludes this paper.

II. Multi-Fidelity Kriging

The key idea of any multi-fidelity surrogate model is to map the trend of the unknown function underlying the intensively sampled low-fidelity (LF) data to the less intensively sampled high-fidelity (HF) data. For the present work a hybrid bridge function approach adopted from Han et al.\textsuperscript{25} was implemented and has been presented in previous work.\textsuperscript{41} The relationship between the high- and low-fidelity surrogate model (surrogate values indicated by a hat) in any location $x$ is expressed as:

\[
\hat{y}_{HF}(x) = \hat{\phi}(x)\hat{y}_{LF}(x) + \hat{\gamma}(x)
\]

where $\hat{\gamma}(x)$ is an additive bridge function and $\hat{\phi}(x) = f^T(x)\hat{\rho}$ is a low-order polynomial with $q + 1$ basis functions $f^T(x) = [1, f_1(x), \ldots, f_q(x)]$ with corresponding coefficients $\hat{\rho} = [\hat{\rho}_0, \hat{\rho}_1, \ldots, \hat{\rho}_q]^T$. The implemented framework supports an arbitrary number of fidelity levels via an approach similar to a multi-grid strategy.
The construction of a multi-fidelity surrogate (MFS) model is accomplished via the following four steps:

1. Build kriging model for lowest fidelity data, \( \hat{y}_{LF_1} \), using \( N_{LF_1} \) lowest fidelity training points

2. Build another kriging model for the additive bridge function, \( \gamma_2 \), connecting to next fidelity level, where \( \gamma_2(x) = y_{LF_2}(x) - \phi_2(x)\hat{y}_{LF_1}(x) \) in \( N_{LF_2} \) next fidelity level training points

3. Compute optimal \( \hat{\rho}_2 \) during the maximum likelihood estimation updates for \( \gamma_2 \) which yields \( \hat{y}_{LF_2}(x) = \phi_2(x)\hat{y}_{LF_1}(x) + \gamma_2(x) \)

4. If \( \gamma_2 \neq 0 \) stop, otherwise repeat steps 2 and 3 until highest fidelity level has been reached

Intuition suggests and numerical experimentation confirmed that LF training point locations should contain the HF locations as a true subset to ensure that no additional error through \( \hat{y} \) is introduced when computing \( \gamma(x) = y_{HF}(x) - \phi(x)\hat{y}_{LF}(x) \) in all \( N_{HF} \) high-fidelity training point locations. The remaining \( (N_{LF} - N_{HF}) \) LF points are picked via Latin hypercube sampling (LHS) subject to a distance constraint to the existing HF points. Also, whenever a HF point is added via the dynamic training point algorithm the corresponding LF point is added to the set as well.

### III. Multi-fidelity Sparse Polynomial Chaos Expansions

Usually, multi-fidelity PCE implementations take a multi-step approach, separating the estimation of corrective functions from the fitting of the model of interest.\(^{42-44}\) The all-at-once approach developed by Bryson and Rumpfkeil,\(^{13,21}\) which is employed here, differs in that only a single fitting is performed, simultaneously determining additive and multiplicative corrections to the low-fidelity data to best approximate the high-fidelity function in a least-squares sense. The all-at-once approach can also be augmented with the addition of gradient and Hessian information for both high- and low-fidelity data if it is available.

In a traditional PCE, a high-fidelity function is approximated by the series

\[
\hat{y}_{HF}(x) = \sum_{k=0}^{P} \hat{\beta}_k \Psi_k(x)
\]

(2)

where \( \hat{\beta}_k \) are polynomial coefficients and \( \Psi_k(x) \) are the selected multidimensional bases. For the inclusion of multi-fidelity data in the model fitting, it is assumed that the high-fidelity function is matched by multiplicative and additive corrections to the low-fidelity model,

\[
\hat{y}_{HF}(x) = \hat{y}_{LF}(x) + \alpha(x) \hat{y}_{LF}(x) + \delta(x).
\]

(3)

where the corrective terms, \( \alpha(x) \) and \( \delta(x) \), themselves can be represented as PCEs as well. Estimated values of the PCE coefficients (\( \hat{\beta}_k, \hat{\alpha}_k, \) and \( \hat{\delta}_k \)) can then be determined by a singular value decomposition, which minimizes the errors relative to the training data in a least-squares sense and is relatively robust to ill-conditioned systems.

#### III.A. Training Point Selection

Using a regression procedure to solve for the PCE coefficients via non-intrusive point collocation, the selection of training points is decoupled from the underlying basis functions. However, thoughtful selection may help to improve the convergence of the surrogate model. Training points and basis functions of the PCE may be selected to correspond to the underlying probability distributions of the input parameters. For the considered benchmark problems in this paper the design parameters are assumed to be uniformly distributed, corresponding to Legendre basis polynomials. The training points are therefore selected to be Gauss-Patterson knots and each grid level is a subset of the next higher level grid. This nested property is desired to enable reusing data in an adaptive process. Structured grids may be built upon tensor products of one-dimensional abscissae, but the number of points grows exponentially with the number of dimensions and grid level. To ameliorate the cost of evaluating the responses on such a mesh, Smolyak sparse grids with a slow-exponential growth rule are generated using Burkardt’s sparse grid mixed growth anisotropic rules library,\(^{45}\) and the number of resulting training points for several design space dimensions and grid levels are
given in Table 1. In the multi-fidelity cases, a LF grid (of which the HF grid is fully a subset) is specified to be at least one grid level higher.

Table 1. Number of points for several grid levels and dimensions of slow-growth, sparse, Gauss-Patterson grids.

<table>
<thead>
<tr>
<th>Grid Level</th>
<th>Dimension, $D$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>10</th>
<th>20</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>5</td>
<td>7</td>
<td>9</td>
<td>11</td>
<td>13</td>
<td>21</td>
<td>41</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>9</td>
<td>19</td>
<td>33</td>
<td>51</td>
<td>73</td>
<td>201</td>
<td>401</td>
<td>801</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>17</td>
<td>39</td>
<td>81</td>
<td>151</td>
<td>257</td>
<td>1201</td>
<td>3201</td>
<td>8241</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>33</td>
<td>87</td>
<td>193</td>
<td>391</td>
<td>737</td>
<td>5281</td>
<td>90561</td>
<td>1557121</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>33</td>
<td>135</td>
<td>385</td>
<td>903</td>
<td>1889</td>
<td>19105</td>
<td>641409</td>
<td>huge</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>65</td>
<td>207</td>
<td>641</td>
<td>1743</td>
<td>4161</td>
<td>60225</td>
<td>huge</td>
<td>huge</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>97</td>
<td>399</td>
<td>1217</td>
<td>3343</td>
<td>8481</td>
<td>169185</td>
<td>huge</td>
<td>huge</td>
</tr>
</tbody>
</table>

III.B. Compressed Sensing

There exists a wide variety of methods for sparse recovery of a signal $\hat{\beta}$ from a set of incomplete measurements. Knowing (or hoping) that the original signal is sparse, and having obtained the measurement vector $b$, one can attempt to recover $\hat{\beta}$ by solving an optimization problem of the form

$$\min_{\hat{\beta}} 0.5|A\hat{\beta} - b|^2_2 + \lambda||\hat{\beta}||_1 \quad (4)$$

where $A$ is the regression matrix and $\lambda$ is a penalty coefficient. Here, the LASSO (least absolute shrinkage and selection operator) algorithm via the software library mlpack is employed since it can provably solve this optimization problem. Note, that selecting a proper value for $\lambda$ in equation (4) is very important for an accurate computation of the sparse PCE coefficients. If $\lambda$ is too large, then the reconstructed signal may not be accurate enough, while very small values of $\lambda$ may result in over-fitting or a less sparse solution.

Following Salehi et al., either a true root-mean-square error (RMSE) or the leave-one-out cross-validation error has been used in previous work and is employed here to determine an optimal $\hat{\lambda}$. In order to find it, the LASSO is solved for multiple values of $\lambda$, and the associated error is calculated for each case. Then, $\hat{\lambda}$ is simply set to the value of $\lambda$ for which the error is the smallest.

IV. Agglomeration of Locally Optimized Surrogates (ALOS)

Following Liem et al., an explicit mixture-of-experts approach is employed here for a heterogeneous non-polynomial test function. The problem domain is first partitioned into $K$ subregions via a Gaussian mixture model (GMM) unsupervised learning algorithm, which is followed by local expert (surrogate model) training in each subregion, which is likely more homogeneous than the entire domain. The local predictions can then be combined probabilistically to yield the agglomerated final prediction, $\hat{y}(x)$, in any location $x$:

$$\hat{y}(x) = \sum_{k=1}^{K} \Pi_k(x)\hat{y}_k(x) \quad (5)$$

where $\hat{y}_k(x), k = 1, \ldots, K$ are the local surrogate model predictions (ie. the local experts) and $\Pi_k(x)$ is the mixing proportion which is given by a modified softmax function. The agglomeration of locally optimized surrogates (ALOS) framework can be summarized as follows:

1. Perform unsupervised learning using a GMM to initially cluster the available training data. The program EMMIX (Expectation-Maximization-based MIXture analysis) is employed here. The user must decide on the clustering criterion (default is function values). The number of clusters is maximized by the algorithm such that at least a user defined minimum number of training points is in each cluster. The algorithm can check for “islands” which may have been created and remove them automatically.
by reassigning all training points in the island to the largest touching cluster.

2. Build a separate local surrogate model within each cluster, \( \hat{y}_k(x) \), \( k = 1, \ldots, K \), and compute the corresponding mixing proportion, \( \Pi_k(x) \). The clusters can either have no overlap at all (i.e., they are a partition of unity) or a user-defined number of training points from each neighboring cluster can be added to the local training set to promote continuity between clusters.

3. Compute the agglomerated estimation, \( \hat{y}(x) \), using equation (5).

4. Add additional training points to the clusters and return to step 2 until a computational budget is exhausted or convergence.

V. Benchmark Problems

The following benchmark problems are proposed and discussed in more detail in the following subsections.

1. A polynomial test function (any dimension)
2. A non-polynomial test function (any dimension)
3. Heterogeneous non-polynomial test function (one-, two-, and three-dimensional)
4. Coupled Spring-Mass-System leading to a linear system of ODEs (any dimension)

V.A. Analytic Test Functions

To test the surrogate models described in the previous sections, two analytic functions are employed in arbitrary dimensions over the domain \(-2 \leq x_i \leq 2\) for \( i = 1, \ldots, D \). The first one is the well-known Rosenbrock function,

\[
r(x) = \sum_{i=1}^{D-1} (1 - x_i)^2 + 100 (x_{i+1} - x_i^2)^2.
\]

Globally it is easy to approximate the overall quartic shape of this function, but locally it is difficult to capture the gently sloped valleys leading to a global minimum. The second function is a sum of exponential-sine products, which is not a polynomial and thus cannot be fully represented by PCEs,

\[
e(x) = \frac{1}{D} \sum_{i=1}^{D} \exp(-0.1x_i) \sin\left(\frac{\pi}{2}x_i\right).
\]

The LF version of the Rosenbrock function is a transformation of the HF function value by linear additive and multiplicative factors,

\[
l(x) = \frac{r(x) - 4.0 - \sum_{i=1}^{D} 0.5x_i}{10.0 + \sum_{i=1}^{D} 0.25x_i}.
\]

Thus, linear hybrid corrections to the LF data will fully recover the HF function. For the sum of exponential-sine products an LF expression is taken to be

\[
l(x) = \frac{1}{D} \sum_{i=1}^{D} \sin\left(\frac{\pi}{2}x_i\right).
\]

which mimics the global behavior of the HF function but lacks some detail. Both high-fidelity functions are shown in two dimensions in Figure 1.

Heterogeneous non-polynomial analytic functions defined on unit cubes in one, two, and three dimensions are considered as benchmark applications for the agglomeration of locally optimized surrogates (ALOS). The one- and two-dimensional high-fidelity functions are taken from Clark and Bae and are given by

\[
f(x) = \sin[30(x - 0.9)^4] \cos[2(x - 0.9)] + (x - 0.9)/2
\]
and

\[ g(x, y) = \sin[21(x - 0.9)^4] \cos[2(x - 0.9)] + (x - 0.7)/2 + 2y^2 \sin[xy] \]  

(11)

An extension to three dimensions is given by

\[ h(x, y, z) = \sin[21(x - 0.9)^4] \cos[2(x - 0.9)] + (x - 0.7)/2 + 2y^2 \sin[xy] + 3z^3 \sin[xyz] \]  

(12)

Low-fidelity versions are obtained by using linear additive and multiplicative bridge functions:

\[ f_l(x) = (f - 1.0 + x)/(1.0 + 0.25x) \]  

(13)

\[ g_l(x, y) = (g - 2.0 + x + y)/(5.0 + 0.25x + 0.5y) \]  

(14)

\[ h_l(x, y, z) = (h - 2.0 + x + y + z)/(5.0 + 0.25x + 0.5y - 0.75z) \]  

(15)

The one- and two-dimensional functions are displayed together with their low-fidelity approximations in Figure 2.
V.B. Coupled Spring-Mass-System

Three masses are attached to each other by four springs as shown in Figure 3.

![Figure 3. Three masses connected by springs. The masses slide along a frictionless horizontal surface.](image)

The analysis uses the following constants, variables and assumptions.

**Mass Constants** The masses \( m_1, m_2, m_3 \) are assumed to be point masses concentrated at their center of gravity.

**Spring Constants** The mass of each spring is negligible. The springs operate according to Hooke's law and the constants \( k_1, k_2, k_3, k_4 \) denote the Hooke's constants. The springs restore after compression and extension.

**Position Variables** The symbols \( x_1(t), x_2(t), x_3(t) \) denote the mass positions along the horizontal surface, measured from their equilibrium positions, positive right and negative left.

**Fixed Ends** The first and last spring are attached to fixed walls.

The equations of motion are given by

\[
\begin{align*}
    m_1 \ddot{x}_1(t) &= -k_1 x_1(t) + k_2 [x_2(t) - x_1(t)] \\
    m_2 \ddot{x}_2(t) &= -k_2 [x_2(t) - x_1(t)] + k_3 [x_3(t) - x_2(t)] \\
    m_3 \ddot{x}_3(t) &= -k_3 [x_3(t) - x_2(t)] - k_4 x_3(t)
\end{align*}
\]  

These are justified in the case of all positive variables by observing that the first three springs are elongated by \( x_1, x_2 - x_1, \) and \( x_3 - x_2, \) respectively. The last spring is compressed by \( x_3, \) which accounts for the minus sign. Another way to justify the equations is through mirror-image symmetry: interchange \( k_1 \) with \( k_4, \) \( k_2 \) with \( k_3, \) and \( x_1 \) with \( x_3, \) then the middle equation should be unchanged and the last equation should become the first equation and vice versa.

The equations (16) can be written as a second-order matrix-vector system

\[
M \ddot{x}(t) = K x(t)
\]  

where the displacement \( x, \) mass matrix \( M \) and stiffness matrix \( K \) are defined by the following formulae

\[
x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \quad M = \begin{pmatrix} m_1 & 0 & 0 \\ 0 & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}, \quad K = \begin{pmatrix} -k_1 - k_2 & k_2 & 0 \\ k_2 & -k_2 - k_3 & k_3 \\ 0 & k_3 & -k_3 - k_4 \end{pmatrix}
\]

This is a constant-coefficient homogeneous system of second-order ODEs the solution of which is given by

\[
x(t) = \sum_{i=1}^{3} [a_i \cos(\omega_i t) + b_i \sin(\omega_i t)] v_i
\]  

where \( \omega_i = \sqrt{-\lambda_i} \) and \( \lambda_i \) are the eigenvalues of the matrix \( M^{-1} K \) and \( v_i \) are the corresponding eigenvectors. The constants \( a_i \) and \( b_i \) are determined by the initial conditions \( x(t = 0) = x_0 \) and \( \dot{x}(t = 0) = \dot{x}_0 \)
**Example Solution**

As a small numerical example let \( m_1 = m_2 = 1, k_1 = k_3 = 4 \) and \( k_2 = 2 \) as well as \( x_0 = (1 \ 0)^T \) and \( \dot{x}_0 = (0 \ 0)^T \) then

\[
M^{-1}K = \begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}
\]

with real eigenvalues \( \lambda_1 = -1 \) and \( \lambda_2 = -3 \) and corresponding eigenvectors

\[
v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}
\]

Since the initial velocity is zero we have \( b_1 = b_2 = 0 \) and using \( x(t = 0) = x_0 \) in equation (18) yields \( a_1 = a_2 = 0.5 \) such that

\[
x_1(t) = 0.5 \cos(t) + 0.5 \cos(\sqrt{3}t) \quad \text{and} \quad x_2(t) = 0.5 \cos(t) - 0.5 \cos(\sqrt{3}t)
\]

These solutions are plotted in Figure 4 for \( 0 \leq t \leq 30 \).

**Proposed Benchmark Problem**

Converting equation (17) into a system of first-order ODEs and using the fourth-order accurate Runge-Kutta time-marching method yields a multi-fidelity analysis problem by varying the time-step size \( \Delta t \) as shown in Figure 5. Note that the HF analysis with \( \Delta t = 0.01 \) is virtually indistinguishable from the analytical solution given by equation (19) while the LF analysis with \( \Delta t = 0.6 \) is exhibiting the correct trends but is somewhat inaccurate.
Figure 5. Numerical solution for $x_1(t)$ of the example problem for $0 \leq t \leq 6$ using $\Delta t = 0.01$ labeled as high-fidelity (HF) and $\Delta t = 0.6$ labeled as low-fidelity (LF).

Treating the two masses as independent input variables with $1 \leq m_1, m_2 \leq 4$ while $k_1 = k_3 = 4$ and $k_2 = 2$ and computing $x_1(t = 6)$ yields the two-dimensional design space shown in Figure 6.

Figure 6. $x_1(t = 6)$ when $1 \leq m_1, m_2 \leq 4$ with HF results shown in red and LF results shown in blue.

Similarly, treating the two spring constants as independent input variables with $1 \leq k_1, k_2 \leq 4$ while $m_1 = m_2 = 1$ yields the two-dimensional design space shown in Figure 7. One can see that in both cases the lower fidelity trends match the high-fidelity ones which is required for the use of a multi-fidelity approach.
Here are some suggestions for multi-fidelity surrogate model benchmark problems:

1. Build the two-dimensional surrogate models for Figures 6 or 7 and compute the root-mean-square error (RMSE) compared to HF data on equispaced Cartesian meshes using $101^2 = 10,201$ nodes

2. Find the HF global minimum in Figure 7 located at about $(2.468, 2.193)$. Note that the LF global minimum does not quite coincide with the HF one.

3. Various uncertainty quantification problems can be devised such as

   (a) Treat the independent input variables as aleatory with a known probability distribution and compute the mean and variance of $x_1(t = 6)$
   (b) Treat the initial conditions as aleatory with a known probability distribution and compute the mean and variance of $x_1(t = 6)$
   (c) Combine the uncertainties from (a) and (b)
   (d) Treat some or all uncertainties as epistemic

4. The dimensionality of the design space can be easily increased by adding more masses and springs to the system. More output functions of interest are given by the position variables of the other masses.

5. A dimension reduction strategy can be implemented by replacing multiple springs and masses with fewer springs and masses.

6. Another interesting aspect, that could be exploited in this problem, is the existence of beat frequencies that occur if two of the normal mode frequencies get close to each other.
VI. Results

In this section, the root-mean-square error (RMSE) is used for comparison purposes and is computed on equispaced Cartesian meshes using 1001 nodes in 1D, $101^2 = 10,201$ nodes in 2D, $51^3 = 132,651$ nodes in 3D, $8^6 = 262,144$ nodes in 6D, and $3^{10} = 59,049$ nodes in 10D. Note that lines are plotted between markers in the result figures as a visual aide, but do not necessarily reflect performance at intermediate points.

VI.A. Rosenbrock Function

The Rosenbrock function serves as a verification case for the correct implementation of the MFSPCE method and results are provided in Figure 8 for two and ten dimensions.

![Figure 8](image_url)

Figure 8. Verification of multi-fidelity PCE implementation on Rosenbrock function in two and ten dimensions.

Once sufficient training data are gathered to construct a fourth-order polynomial expansion, the surrogate model should match the truth function exactly with the error dropping accordingly. Imposing an oversampling ratio of two, the standard (full order) PCE surrogate model with high-fidelity training data alone (HF) requires 30 training points to match the exact function in two dimensions and 2002 points in ten dimensions. For the multi-fidelity (MF) models some of these points may be LF. The number of LF training points in two dimensions is four levels higher (i.e., when using 17 HF points, 97 LF points are used) and in ten dimensions it is one level higher (e.g., 201 HF are combined with 1201 LF points). The shown results are in agreement with the fundamental expectations, demonstrating the correctness of the implementation.

For the sparse polynomial chaos expansion models, a fourth-order polynomial is requested (and a first-order one for the additive and multiplicative corrections), and the expectation is that the LASSO finds the pertinent zero and non-zero PCE coefficients. As displayed in Figure 8, this works well for the high-fidelity sparse (HFS) as well as multi-fidelity sparse (MFS) polynomial chaos expansion models yielding an RMSE...
that is lower than the corresponding standard PCE. In particular, the two-dimensional MFS model is orders of magnitude better. It should be noted that once the number of training points is large enough to obtain the exact fourth-order polynomial, the sparse models are not quite as good as the full order PCEs (compare solid to dashed lines) due to the extra non-zero term $\lambda ||\hat{\beta}||_1$ in equation (4).

Figure 9 shows the performance of the kriging model using the same scales as the results of the MFSPCE. One can infer that in two dimensions adding either gradient information (FG) or 40 low-fidelity training points (40 LF) is beneficial while adding both yields no extra gains. In ten dimensions it is computationally too expensive to build the kriging model with gradient information and there is no substantial improvement through adding LF training data. In any case, the performance of the multi-fidelity kriging is much worse than the MFSPCE which is to be expected since the underlying Rosenbrock function is a polynomial.

![Figure 9. Multi-fidelity kriging results for Rosenbrock function in two and ten dimensions.](image)

VI.B. Exponential-sine Function

For the sum of exponential-sine products, the multi-fidelity surrogate model essentially needs the freedom to multiply the LF function by the polynomial expansion of the exponential function and subtract off the cross-dimensional terms. While it is still expected that multi-fidelity data will be helpful, it will be not as powerful as in the Rosenbrock function example since the required bridge function is not a simple polynomial itself. The number of LF training points for the MFSPCE models in the considered two and six dimensions is always three and one level higher, respectively. For the sparse polynomial chaos expansion models, many different orders for the polynomials are combined with various values for $\lambda$ to find the combination with the lowest RMSE. The resulting best values are shown in Tables 2 and 3 for the two- and six-dimensional case, respectively.

![Figure 9. Multi-fidelity kriging results for Rosenbrock function in two and ten dimensions.](image)

<table>
<thead>
<tr>
<th># of HF points</th>
<th># of LF points</th>
<th>$\lambda$</th>
<th>Order, $P$</th>
<th>Additive, $R$</th>
<th>Multiplicative, $Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>$10^{-6}$</td>
<td>4</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>$10^{-3}$</td>
<td>5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>$10^{-3}$</td>
<td>6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>33</td>
<td>0</td>
<td>$10^{-6}$</td>
<td>7</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>65</td>
<td>0</td>
<td>$10^{-6}$</td>
<td>10</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>$10^{-3}$</td>
<td>9</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>33</td>
<td>$10^{-3}$</td>
<td>10</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>17</td>
<td>65</td>
<td>$10^{-6}$</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>33</td>
<td>97</td>
<td>$10^{-6}$</td>
<td>9</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>65</td>
<td>161</td>
<td>$10^{-8}$</td>
<td>9</td>
<td>6</td>
<td>6</td>
</tr>
</tbody>
</table>
Table 3. Best order and \( \lambda \) values for six-dimensional exponential-sine function.

<table>
<thead>
<tr>
<th># of HF points</th>
<th># of LF points</th>
<th>( \lambda )</th>
<th>Order, ( P )</th>
<th>Additive, ( R )</th>
<th>Multiplicative, ( Q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>0</td>
<td>( 10^{-6} )</td>
<td>4</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>73</td>
<td>0</td>
<td>( 10^{-3} )</td>
<td>4</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>257</td>
<td>0</td>
<td>( 10^{-3} )</td>
<td>5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>737</td>
<td>0</td>
<td>( 10^{-6} )</td>
<td>6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>13</td>
<td>73</td>
<td>( 10^{-4} )</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>73</td>
<td>257</td>
<td>( 10^{-2} )</td>
<td>8</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>257</td>
<td>737</td>
<td>( 10^{-6} )</td>
<td>8</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>737</td>
<td>1889</td>
<td>( 10^{-6} )</td>
<td>9</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

The results for the exponential-sine function with these settings are plotted in Figure 10 for two and six dimensions. One can infer, generally speaking, that the sparse PCE surrogate models outperform the full order ones (compare solid to dashed lines), and the multi-fidelity models are better than the mono-fidelity ones (compare red to black lines). In six dimensions this is especially true for 73 HF training points (and 257 LF, if requested) and in two dimensions for 33 HF training points (and 97 LF ones).

![Figure 10. RMSE comparisons for multidimensional exponential-sine function using MFSPCE.](image)

Figure 11 shows the performance of the kriging model using the same scaling as the presented MFSPCE results. One can again infer that in two dimensions adding either gradient information (FG) or thirty low-

![Figure 11. Multi-fidelity kriging results for exponential-sine function in two and six dimensions.](image)
fidelity training points (30 LF) is beneficial while adding both yields no extra gains. In six dimensions it is computationally too expensive to build a multi-fidelity kriging model with gradient information and there is no substantial improvement through the addition of LF training data, however, adding gradient information is beneficial. Overall, the performance of the multi-fidelity kriging is comparable or arguably slightly better than the one of MFSPCE.

VI.C. ALOS Function

One-dimensional ALOS Function

In Figure 12 the performances of agglomerated locally optimized surrogates (ALOS) and a global kriging as well as MFSPCE surrogate for the one-dimensional function are compared.

![Figure 12](image-url)

Figure 12. RMSE between the one-dimensional truth model and ALOS with 2 clusters as well as a global surrogate both enhanced with low-fidelity data as a function of number of high-fidelity training points.

One can observe that the ALOS approach outperforms the global approach for both function value only (F) and function and gradient values (FG). The global kriging model errors tend to be larger due to the required compromise in the distance weight or length scale, $\theta$, in the kriging model which has to be a single value in each dimension. When the modeling is distributed to local experts one can disregard the correlation between samples from different subregions. Moreover, each local expert is free to select the best model parameters to better reflect the characteristics of the underlying function in its subregion (e.g., different distance weights or covariance functions for each local kriging model). For the ALOS PCE the use of fewer training points and the smaller domain sizes for the local expert training reduce fluctuations and thus errors.

Figure 12 also shows the performance when the global and ALOS models are enhanced with low-fidelity data. One can infer that the LF data tends to reduce the errors especially in the beginning (compare dashed and solid lines in same color). The poorer performance at the end is likely due to over-fitting.
Plots of the initial kriging models are shown in Figure 13 and one can observe the larger fluctuation of the global model in the right half of the domain which is due to the required compromise in \( \theta \). The optimal distance weight for the first local model (left half) is \( \theta = 7.43 \) whereas for the second (right half) it is \( \theta = 2.41 \) where a lower \( \theta \) implies a stronger correlation between the training points. In contrast, for the global kriging the optimal \( \theta \) is 7.20 which shows the required compromise and yields a seven times larger RMSE as can be inferred from Figure 12 and visually seen in Figure 13.

**Two-dimensional ALOS Function**

In Figure 14 the performances of ALOS with two clusters and a global kriging surrogate as well as MFSPCE surrogate for the two-dimensional function are compared.

One can infer that ALOS outperforms the global approach in the beginning and end for function value only (F) whereas it is worse than the global approach for function and gradient values (FG) in the beginning but then does much better than the global approach. Figure 14 also shows the performance with enhancement through low-fidelity data. The LF data tends to reduce the errors for all numbers of high-fidelity training points considered (compare dashed and solid lines in same color).

Figure 15 shows the domains of the two clusters and the training point distribution at the beginning and the end of the kriging simulation. At the end one can observe that the dynamic training point algorithm put the points especially in the second (blue) cluster (see also Figure 17) which also features the largest \( \theta \) values as shown in Table 4. One can also observe a smoother transition and thus less agglomerated error between the two clusters compared to the beginning.

The local kriging and gradient-enhanced kriging (GEK) models have different optimum length scales \( \theta = [\theta_x, \theta_y] \) in the partitioned input space as the global kriging and GEK as shown in Table 4, suggesting that the ALOS approach has the potential to be better at modeling the different characteristics in the function profile. The difference can be especially seen in the more non-stationary \( x \)-direction where the global kriging’s \( \theta \)-value of 2.55 is a compromise between the blue cluster’s 4.20 and red cluster’s 0.59 where a lower \( \theta \) implies a stronger correlation between the training points. A consequence of this is shown in Figure 16 where plots of the initial models are displayed. The global kriging model exhibits additional fluctuation in the relatively smooth right half of the figure whereas ALOS kriging is much more accurate there. However, ALOS exhibits additional fluctuations in the overlap region due to the mismatch of the two local models as shown in the left of Figure 15.
Figure 14. RMSE between the two-dimensional truth model and ALOS with 2 clusters as well as a global surrogate both enhanced with low-fidelity data as a function of number of high-fidelity training points.

(a) MFSPCE
(b) ALOS Kriging
(c) MFLOS Kriging

Figure 15. Plot of $\Pi_k(x)$ using the 25 initial points (left) and final 63 points (right). Different colors correspond to different clusters. The highest color intensity within each cluster corresponds to $\Pi_k(x) = 1$. 
<table>
<thead>
<tr>
<th>Number of clusters</th>
<th>Length scales, ( \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>kriging</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>[2.55, 0.27]</td>
</tr>
<tr>
<td>2</td>
<td>[0.59, 0.21], [4.20, 0.34]</td>
</tr>
<tr>
<td>GEK</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>[0.55, 0.04]</td>
</tr>
<tr>
<td>2</td>
<td>[0.32, 0.06], [0.59, 0.04]</td>
</tr>
</tbody>
</table>

Table 4. Optimum length scales for kriging and GEK using 1 and 2 clusters in two dimensions using 25 training points.

Figure 16. Plot of truth model (red) as well as the global kriging (white) and ALOS with 2 clusters (blue) using the 25 initial points.

A plot of the final kriging models after 38 points have been dynamically added (for a total of 63) is shown in Figure 17 where an excellent agreement between ALOS and the truth model can be visually inferred.

Figure 17. Plot of truth model (white) as well as the two local surrogate models (same color code as in Figure 15) using 63 points.
Three-dimensional ALOS Function

In Figure 18 the performances of ALOS with two clusters and a global kriging surrogate as well as MFSPCE surrogate for the three-dimensional function are compared.

![Graphs showing RMSE against number of HF training points for different models: MFSPCE, ALOS Kriging, and MFLOS Kriging.](image)

(a) MFSPCE  
(b) ALOS Kriging  
(c) MFLOS Kriging

Figure 18. RMSE between the three-dimensional truth model and ALOS with 2 clusters as well as a global surrogate both enhanced with low-fidelity data as a function of number of high-fidelity training points.

One can infer that ALOS somewhat outperforms the global approach. It also shows the performance with enhancement through low-fidelity data where the LF data tends to reduce the errors for up to 100 HF training points (compare dashed and solid lines in same color) at which point over-fitting is likely occurring.

The local kriging and gradient-enhanced kriging (GEK) models have again different optimum length scales \( \theta = [\theta_x, \theta_y, \theta_z] \) in the partitioned input space as the global kriging and GEK as shown in Table 5, again suggesting that the ALOS approach has the potential to be better at modeling the different characteristics in the function profile.

<table>
<thead>
<tr>
<th>Number of clusters</th>
<th>Length scales, ( \theta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>kriging</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>[0.27, 0.18, 0.28]</td>
</tr>
<tr>
<td>2</td>
<td>[0.09, 0.10, 0.19], [0.24, 0.20, 0.24]</td>
</tr>
<tr>
<td>GEK</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>[0.84, 0.51, 0.33]</td>
</tr>
<tr>
<td>2</td>
<td>[0.81, 0.44, 0.26], [0.51, 0.29, 0.18]</td>
</tr>
</tbody>
</table>

Table 5. Optimum length scales for kriging and GEK using 1 and 2 clusters in three dimensions using 64 training points.
Figure 19 shows the domains of the two clusters and the training point distribution at the end of the kriging simulation. One can observe that the dynamic training point algorithm placed new training points especially in the second (red) cluster (which also features the largest $\theta$ values as shown in Table 5).

VI.D. Coupled Spring-Mass-System

For the MFSPCE models of the coupled spring-mass-system, many different orders for the polynomials are combined with various values for $\lambda$ to find the combination with the lowest RMSE. The resulting best values are shown in Table 6.

<table>
<thead>
<tr>
<th># of HF points</th>
<th># of LF points</th>
<th>$\lambda$</th>
<th>Order, $P$</th>
<th>Additive, $R$</th>
<th>Multiplicative, $Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0</td>
<td>$10^{-1}$</td>
<td>6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>$10^{-3}$</td>
<td>9</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>$10^{-2}$</td>
<td>9</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>33</td>
<td>0</td>
<td>$10^{-4}$</td>
<td>6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>65</td>
<td>0</td>
<td>$10^{-8}$</td>
<td>7</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>97</td>
<td>0</td>
<td>$10^{-8}$</td>
<td>9</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>161</td>
<td>0</td>
<td>$10^{-8}$</td>
<td>11</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>33</td>
<td>$10^{-4}$</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>33</td>
<td>$10^{-4}$</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>65</td>
<td>$10^{-7}$</td>
<td>10</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>33</td>
<td>97</td>
<td>$10^{-8}$</td>
<td>10</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>65</td>
<td>161</td>
<td>$10^{-8}$</td>
<td>11</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>97</td>
<td>161</td>
<td>$10^{-8}$</td>
<td>11</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>161</td>
<td>257</td>
<td>$10^{-8}$</td>
<td>11</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 6. Best order and $\lambda$ values for the two-dimensional coupled spring-mass-system using MFSPCE.
The results for these settings are plotted on the left in Figure 20 whereas the performance of the kriging model is shown on the right using the same scaling in both figures.

![Figure 20. RMSE comparisons for two-dimensional mass-spring system.](image)

In general, one can infer that the sparse PCE surrogate models outperform the full order ones (compare solid to dashed lines), and the multi-fidelity models are better than the mono-fidelity ones (compare red to black lines) for both the MFSPCE and MF kriging models. Overall, the performance of the multi-fidelity kriging is comparable to the one of MFSPCE.

In addition to the two-dimensional case a four-dimensional one where both spring constants and both masses are varied between one and four is considered here as well. The RMSE is calculated on an equispaced Cartesian mesh using $12^4 = 20,736$ nodes. Again, many different orders for the MFSPCE are combined with various values for $\lambda$ to find the combination with the lowest RMSE. The resulting best values are given in Table 7 and the results are shown in Figure 21.

<table>
<thead>
<tr>
<th># of HF points</th>
<th># of LF points</th>
<th>$\lambda$</th>
<th>Order, $P$</th>
<th>Additive, $R$</th>
<th>Multiplicative, $Q$</th>
</tr>
</thead>
<tbody>
<tr>
<td>33</td>
<td>0</td>
<td>$10^{-1}$</td>
<td>5</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>81</td>
<td>0</td>
<td>$10^{-1}$</td>
<td>4</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>193</td>
<td>0</td>
<td>$10^{-1}$</td>
<td>7</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>385</td>
<td>0</td>
<td>$10^{-3}$</td>
<td>7</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>33</td>
<td>81</td>
<td>$10^{-2}$</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>81</td>
<td>193</td>
<td>$10^{-2}$</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>193</td>
<td>385</td>
<td>$10^{-3}$</td>
<td>7</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>385</td>
<td>641</td>
<td>$10^{-3}$</td>
<td>9</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7. Best order and $\lambda$ values for the four-dimensional coupled spring-mass-system using MFSPCE.

Once again, one can infer that the sparse PCE surrogate models outperform the full order ones (compare solid to dashed lines), and the multi-fidelity models are better than the mono-fidelity ones (compare red to black lines) for both the MFSPCE and MF kriging models. Overall, the performance of the multi-fidelity kriging is somewhat better than MFSPCE.

VII. Conclusions

The goal of this work is to build highly accurate surrogate models for applications in design space exploration and optimization as well as uncertainty quantification at small overall computational cost assuming that obtaining high-fidelity training point information is the dominant cost factor. The employed strategy is to enhance surrogate modeling techniques by adaptively selecting training points as well as utilizing derivative information and to employ lower fidelity information as well as locally optimized surrogate models. Four benchmark problems are used to assess the performance of a multi-fidelity sparse polynomial chaos
expansion (MFSPCE) model as well as a multi-fidelity locally optimized surrogate (MFLOS) kriging model. These benchmarks are a polynomial analytic function (any dimension), a non-polynomial analytic function (any dimension), a heterogeneous non-polynomial analytic function (one-, two-, and three-dimensional), and a coupled spring-mass-system leading to a linear system of ODEs (any dimension).

Overall, multi-fidelity models yield more accurate results compared to using high-fidelity data alone especially at the beginning of a simulation when only a handful of high-fidelity training points are available. The sparse PCEs tend to be a factor of two or so better than the full-order counterparts in terms of accuracy as measured by the root-mean-square error (RMSE) compared to a truth model and the agglomeration of locally optimized surrogate (ALOS) approach is also beneficial for a strongly heterogeneous test function.

Acknowledgments

This work was supported by the US Air Force Research Laboratory (AFRL). The second author acknowledges support of the US Air Force Office of Scientific Research (grant 17RQCOR473, Dr. Jean-Luc Cambier, Computational Mathematics Program Officer).

References


