Technical Note

Multi-Fidelity Surrogate Based on Single Linear Regression

Yiming Zhang, Nam-Ho Kim, Chanyoung Park, Raphael T. Haftka
Department of Mechanical and Aerospace Engineering
University of Florida
Gainesville, Florida, 32611
{yimingzhang521, nkim, cy.park, haftka}@ufl.edu

Introduction
Various frameworks have been proposed to predict mechanical system responses by combining data from different fidelities for design optimization and uncertainty quantification as reviewed by Fernández-Godino et al.[1] and Peherstorfer et al. [2]. Among all frameworks, the Bayesian framework based on Gaussian processes [3] has the potential of highest accuracy. However, the Bayesian framework requires optimization for estimating hyper-parameters, and there is a risk of estimating inappropriate hyper-parameters as Kriging surrogate often does, especially in the presence of noisy data. We propose an easy and yet powerful framework for practical design and applications. In this technical note, we revised a heuristic framework [4] which minimizes the prediction errors at high-fidelity samples using optimization. The system behavior (high-fidelity behavior) is approximated by a linear combination of the low-fidelity predictions and a polynomial-based discrepancy function. The key idea is to consider the low-fidelity model as a basis function in the multi-fidelity model with the scale factor as a regression coefficient. The design matrix for least-square estimation consists of both the low-fidelity model and discrepancy function. Then the scale factor and coefficients of the basis functions are obtained simultaneously using linear regression, which guarantees the uniqueness of fitting process. Besides enabling efficient estimation of the parameters, the proposed least-squares multi-fidelity surrogate (LS-MFS) can be applicable to other regression models by simply replacing the design matrix. Therefore, the LS-MFS is expected to be easily applied to various applications such as prediction variance, D-optimal designs, uncertainty propagation [6, 7] and design optimization.

Least-squares Multi-fidelity Surrogate
A heuristic multi-fidelity surrogate (MFS) using polynomial response surface (PRS) has been proposed[4] and demonstrated reasonable effectiveness and robustness. The MFS provides a fit combining a small number of expensive high-fidelity data and less expensive low-fidelity data as a function of design parameters. The MFS is built with two surrogates, \( \hat{f}_s(x) \) and \( \hat{\delta}(x) \), which are the PRS fitted to the low fidelity data and the discrepancy data as

\[
\hat{f}_s(x) = \rho \hat{f}_l(x) + \hat{\delta}(x)
\]

(1)

where the scale factor \( \rho \) and discrepancy function \( \hat{\delta}(x) \) are obtained through optimization as

\[
\min_{\rho, \hat{\delta}(x)} : \left( \hat{\delta}(x_{n}) - d_{n} \right)^{\top} \left( \hat{\delta}(x_{n}) - d_{n} \right) \quad (2)
\]

\[
d_{n} = \rho \hat{f}_l(x_{n}) - y_{n}
\]

(3)

where the high-fidelity data set is denoted as \( \{ x_{n}, y_{n} \} \) including \( n \) samples.

In this technical note, we propose the least-squares estimation of the scale factor \( \rho \) and the coefficients of the discrepancy function. The MFS in Eq. (1) can be expanded with the discrepancy function being represented by \( p \) monomial bases, as
\[
\hat{f}_H(x) = \rho \hat{f}_L(x) + \hat{\delta}(x)
\]

\[
= \rho \hat{f}_L(x) + \sum_{i=1}^{n_H} X_i(x) b_i
\]

where \( X_i(x) \) denotes the \( i \)-th monomial/basis, and \( b_i \) is the coefficient of \( X_i(x) \). In the proposed LS-MFS, the traditional design matrix is augmented by including the low-fidelity model with unknown scale parameter \( \rho \). In least-squares estimation, the relationship between high-fidelity samples and model predictions can be given as

\[
Y = XB + e
\]

where

\[
Y_{n \times 1} = \begin{bmatrix}
Y_H^{(1)} \\
\vdots \\
Y_H^{(n)}
\end{bmatrix},
X_{n \times (p+1)} = \begin{bmatrix}
\hat{f}_L(x_H^{(1)}) & X_1(x_H^{(1)}) & \cdots & X_p(x_H^{(1)}) \\
\vdots & \vdots & \ddots & \vdots \\
\hat{f}_L(x_H^{(n)}) & X_1(x_H^{(n)}) & \cdots & X_p(x_H^{(n)})
\end{bmatrix},
B_{p \times 1} = \begin{bmatrix}
\rho \\
b_1 \\
\vdots \\
b_p
\end{bmatrix},
e_{n \times 1} = \begin{bmatrix}
e_H^{(1)} \\
\vdots \\
e_H^{(n)}
\end{bmatrix}
\]

In the above equation, \( Y \) is the vector of high-fidelity samples, \( X \) is the augmented design matrix, \( B \) is the parameter vector and \( e \) is the vector for residual errors. By augmenting the design matrix, the scale parameter and unknown coefficients of discrepancy functions are estimated simultaneously. In addition, the low-fidelity model is considered as additional basis function. An additional advantage of the proposed LS-MFS is that it is unnecessary to build the surrogate model of the low-fidelity. It only requires to evaluate the low-fidelity model at the same locations with the high-fidelity samples. The unknown parameters in LS-MFS are obtained by using the standard regression technique as

\[
B = (X^T X)^{-1} X^T Y
\]

The scale factor \( \rho \) implies the level of trend similarity between \( \hat{f}_H(x) \) and \( \hat{f}_L(x) \), and plays a critical role to approximate multi-fidelity data. Negative values or extremely large values of \( \rho \) indicates a risky prediction, which are likely to be associated with undesirable low-fidelity models, inappropriate surrogate forms, or inadequate samples. The discrepancy function \( \hat{\delta}(x) \) is likely to be a low-order PRS when \( f_L(x) \) had a similar trend as \( f_H(x) \).

There are nice properties of LS-MFS. Firstly, the LS-MFS could be easily applied to more than two-fidelity models by augmenting the design matrix with multiple low-fidelity models with multiple scale factors, although we showed the LS-MFS for two-fidelity models in this technical note. Secondly, comparing with the heuristic approach, the LS-MFS is more convenient for analytical study of various applications such as the prediction variance, D-optimal design of experiments, uncertainty propagation, and design optimization, while incorporating the effect of the multi-fidelity models.

**Numerical performance of the least-squares multi-fidelity surrogate**

We selected an exponential function [8, 9] with two variables to test the LS-MFS. The high-fidelity model \( f_H(x) \) is given in Eq. (8). A synthetic noise following a normal distribution \( N(0,0.2^2) \) has been added to the high-fidelity samples. We also modified the original low-fidelity model [10] with a larger scale factor and added a quadratic function as in Eq. (9). These variations make the multi-fidelity modeling more challenging. Major settings of the test function are summarized in Table 1. The responses of \( f_H(x) \) (no noise) and \( f_L(x) \) have reasonable spatial complexity as shown in Fig. 1.
\[ f_H(x) = \left[ 1 - \exp\left( -\frac{1}{2x_2} \right) \right] \frac{2300x_1^3 + 1900x_2^2 + 2092x_1 + 60}{100x_1^3 + 500x_2^2 + 4x_1 + 20} \]  

\[ f_L(x) = \frac{1}{8} [f(x_1 + 0.05, x_2 + 0.05) + f(x_1 + 0.05, \max(0, x_2 - 0.05)) + \frac{1}{8} [f(x_1 - 0.05, x_2 + 0.05) + f(x_1 - 0.05, \max(0, x_2 - 0.05))] + \frac{1}{8} (-5x_1 - 7x_2^2) \]  

Table 1. Major settings for the exponential test function

| Input variables \( x_1, x_2 \in [0,1] \) | Range of \( f_H(x) \) | Range of \( f_L(x) \) | Noise for \( f_H(x) \) |
|---------------------------------------|---------------------|---------------------|---------------------|
| \( x_1, x_2 \in [0,1] \) | [1.1804, 13.7692] | [-0.1867, 4.9498] | \( N(0, 0.2^2) \) |

Figure 1. Responses of the test function in 2D space

We generated two sets of samples according to \( 2 \times 2 \) and \( 3 \times 3 \) full factorial designs as shown in Fig. 2(a) and Fig. 2(b). \( y_H \) and \( y_L \) were computed from \( f_H(x) \) and \( f_L(x) \), respectively, at the full factorial designs. Instead of the low-fidelity surrogate \( \hat{f}_L(x) \), the exact low-fidelity model \( f_L(x) \) is used to avoid approximation error in building the low-fidelity surrogate. It is also possible to build \( \hat{f}_L(x) \) based on only low-fidelity data for repeated calls of LS-MFS in practical applications. The LS-MFS was evaluated using relative root-mean-square error (RMSE) for the overall accuracy and maximum error for the worst individual prediction. The relative RMSE and maximum error were obtained based on \( 11 \times 11 \) test grids as shown in Fig. 2(c).

The key parameters and performance of LS-MFS were summarized in Table 2. For the \( 2 \times 2 \) full factorial design, LS-MFS had the relatively small RMSE of 9.00% when approximating the complex test function with noise using a linear discrepancy function. The large maximum relative error, 31.37%, was mainly due to a small function value. While introducing more samples using the \( 3 \times 3 \) design, all the evaluation metrics (e.g. the relative RMSE, the maximum relative/absolute errors) decreased noticeably as in Table 2. PRS fitting for only high-fidelity samples were given in Table 3 as a comparison. For the \( 2 \times 2 \)
design, relative RMSE of PRS fitting was 5 times larger than that of LS-MFS, and the absolute maximum error of PRS fitting was 10 times larger than that of LS-MFS. For the 3×3 design, PRS fitting improved while introducing more samples, but was still significantly inferior to the LS-MFS regarding the specified evaluation metrics.

![Figure 2](image)

Table 2. LS-MFS based on 2×2 and 3×3 full factorial designs

| Model parameters for LS-MFS | Relative RMSE (%) | Maximum relative error (%) | Maximum absolute error |
|-----------------------------|------------------|----------------------------|------------------------|
| 2×2 full factorial design   | ρ=3.40, Linear discrepancy | 9.00% | 31.37% | 0.82 |
| 3×3 full factorial design   | ρ=3.16, Quadratic discrepancy | 1.52% | 8.03% | 0.26 |

Table 3. PRS fitting for only high-fidelity samples based on 2×2 and 3×3 full factorial designs

| Model parameters for LS-MFS | Relative RMSE (%) | Maximum relative error (%) | Maximum absolute error |
|-----------------------------|------------------|----------------------------|------------------------|
| 2×2 full factorial design   | Linear | 43.80% | 92.23% | 8.99 |
| 3×3 full factorial design   | Quadratic | 26.18% | 56.21% | 6.99 |

Acknowledgments

This work was supported by the U.S. Department of Energy, National Nuclear Security Administration, Advanced Simulation and Computing Program, as a Cooperative Agreement under the Predictive Science Academic Alliance Program, under Contract No. DE-NA0002378

References

1. Fernández-Godino, M. G., Park, C., Kim, N.-H., and Hafika, R. T. "Review of multi-fidelity models," arXiv preprint arXiv:1609.07196, 2016.
2. Peherstorfer, B., Willcox, K., and Gunzburger, M. "Survey of multifidelity methods in uncertainty propagation, inference." and optimization. Technical Report TR-16-1. Aerospace Computational Design Laboratory, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA, USA, 2016.
3. Kennedy, M. C., and O'Hagan, A. "Bayesian calibration of computer models," Journal of the Royal Statistical Society: Series B (Statistical Methodology) Vol. 63, No. 3, 2001, pp. 425-464.
4. Zhang, Y., Meeker, J., Schutte, J., Kim, N., and Haftka, R. "On Approaches to Combine Experimental Strength and Simulation with Application to Open-Hole-Tension Configuration," Proceedings of the American Society for Composites: Thirty-First Technical Conference. 2016.

5. Jin, R., Du, X., and Chen, W. "The use of metamodeling techniques for optimization under uncertainty," Structural and Multidisciplinary Optimization Vol. 25, No. 2, 2003, pp. 99-116.

6. Wang, L., Fang, X., Subramaniyan, A., Jothiprasad, G., Gardner, M., Kale, A., Akkaram, S., Beeson, D., Wiggs, G., and Nelson, J. "Challenges in uncertainty, calibration, validation and predictability of engineering analysis models," ASME 2011 Turbo Expo: Turbine Technical Conference and Exposition. American Society of Mechanical Engineers, 2011, pp. 747-758.

7. Sankararaman, S., Ling, Y., and Mahadevan, S. "Uncertainty quantification and model validation of fatigue crack growth prediction," Engineering Fracture Mechanics Vol. 78, No. 7, 2011, pp. 1487-1504.

8. Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments," Journal of the American Statistical Association Vol. 86, No. 416, 1991, pp. 953-963.

9. Bingham, S. S. D. "virtual library of simulation experiments: test functions and datasets." Vol. 2016, http://www.sfu.ca/~ssurjano/index.html, 2016.

10. Xiong, S., Qian, P. Z., and Wu, C. J. "Sequential design and analysis of high-accuracy and low-accuracy computer codes," Technometrics Vol. 55, No. 1, 2013, pp. 37-46.