Jerrad Hampton, October, 2020: Bi-fidelity PCE

This is a casually written document to present a useful estimate with respect to the bi-fidelity PCE method. I am also presenting the work as I see it, to explain that estimate, which may be incorrect, or may be helpful. Traditionally, PCE would be an estimate that did not have a spatial component, only a stochastic component. As such the approximation looks like

$$\hat{u}(\boldsymbol{\xi}) = \sum_{k=1}^{P} c_k \psi_k(\boldsymbol{\xi}), \tag{1}$$

where ξ is of some stochastic dimension d, and $\{\psi_k\}$ are orthogonal polynomials with respect to the distribution of the d-dimensional random vector $\boldsymbol{\Xi}$, and P is identified as a subset of those orthogonal polynomials of lower order, up to e.g. a total-order of p.

The key difference that allows the bi-fidelity PCE to function well, is that the problem now has spatial components, which we will denote by a vector $\boldsymbol{x}$. Now we have a representation of the approximation given by

$$\hat{u}(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{k=1}^{P} c_k \psi_k(\boldsymbol{x}, \boldsymbol{\xi}). \tag{2}$$

We need basis functions, $\psi_k(\boldsymbol{x}, \boldsymbol{\xi})$, which are orthogonal with respect to a measure related to both the spatial and stochastic domains, but also useful for approximating the function. Fortunately, this problem has already been solved for forward UQ simulation with random spatial functions known as KLE. However, there are different correlation functions, and one or more correlation length parameters. That is, the space of potential KLE functions is large.

In what I received, this identification of the KLE was not clearly explained, and these details may matter. In the examples, it seems that this identification is concerned with reducing a stochastic dimension from P to r, where the spatial variance is absorbed into the coefficients, and that is the approach I take here. Specifically we want to change (2) to

$$\hat{u}^{(r)}(\boldsymbol{x}, \boldsymbol{\xi}) = \sum_{k=1}^{r} c_k(\boldsymbol{x}) \eta_k(\boldsymbol{\xi}), \tag{3}$$

for some new functions η_k. This is a rank-reduction compression we are able to address, with the bi-fidelity approximation error. Let $\boldsymbol{\Psi}_{N,P}$ denote the matrix with $\boldsymbol{\Psi}(i,j) = \psi_j(\boldsymbol{x}, \boldsymbol{\xi}_i)$, let $\boldsymbol{c}_{P,1}$ denote a vector of coefficients, and let $\boldsymbol{u}_{N,1}(i) = u(\boldsymbol{x}, \boldsymbol{\xi}_i)$, the evaluation of the solution for the ith realized random variable. The method employed in what I received is to instead have $\boldsymbol{C}_{P,M}$, and $\boldsymbol{U}_{N,M}$ where M is the number of spatial points sampled, giving (with transposes to this problem in what I was sent, which I'll ignore here)

$$\boldsymbol{\Psi}\boldsymbol{C}_{P,M} = \boldsymbol{U}. \tag{4}$$

In this form, the functions ψ_k above have been separated into a stochastic function, which resides in $\boldsymbol{\Psi}$, and a spatial function, which is absorbed into $\boldsymbol{C}$.

We perform the bi-fidelity method to approximate $\boldsymbol{U}$. The theorem we already have bounds,

$$\|\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H\|_F, \tag{5}$$

for $\bar{\boldsymbol{U}}_H$ constructed through the different bi-fidelity algorithm that does not update the coefficient matrix $\boldsymbol{C}$. We use $\hat{\boldsymbol{U}}$ for this version's construction, which does update the coefficient matrix. The method works best if the spatial information is highly compressible, and if it is not (5) will be large, with or without the coefficient update. However, the recomputation of the coefficient matrix $\boldsymbol{C}$ makes this method more adaptive, but also makes that bound from the non-updating algorithm generally looser, and more difficult to apply.

Connected to the bi-fidelity method, we have to identify the r new basis functions η_i, and the matrix,

$$\boldsymbol{\eta}_{N,r} := \boldsymbol{\Psi}_{N,P}\boldsymbol{P}_{P,r}(\boldsymbol{U}^{(r)}), \tag{6}$$

have the name motivated by η in what I was sent. Specifically, η_i represents the function

$$\eta_i(\boldsymbol{\xi}) = \sum_{k=1}^{P} \boldsymbol{\beta}(i,k)\psi_k(\boldsymbol{\xi}), \tag{7}$$

where the affect of the associated projection is given in the coefficients of the orthogonal matrix, $\boldsymbol{\beta}$. These η_i are thus orthogonal, and this orthogonality is critical to the ℓ_2-approximation theory.

This gives the matrix equation which will be used for the final approximation for the PCE itself,

$$\boldsymbol{\eta}_{N_H,r}\boldsymbol{C}_{r,M_H} = \boldsymbol{U}^H_{N_H,M_H}(\boldsymbol{\eta}). \tag{8}$$

That is, $\{u_i\}_{i=1}^{M_H}$ are approximated by $\{\eta_i\}_{i=1}^{r}$, by a set of r coefficients for each of the M_H points. Here, as in what I received, I perform this approximation by M_H independent least squares regressions, which is equivalent to doing the Frobenius regression on the entire matrix. This is the key approximation with high-fidelity samples.

The key low-fidelity computation is the basis identification, specifically how well can the high-fidelity problem be reconstructed by r spatially varying functions. In what I was sent, this has a claimed connection to a KLE, and there are similarities, but this approach does not progress through identifying a covariance kernel, and eigenvalues/functions of it. However, being a KLE or not

does not matter for the approximation here.

To this end, I state our previous bi-fidelity theorem with the matrices that we need. Here we use $\boldsymbol{U}_H^T := \boldsymbol{U}_{N,M_H}$ for $\boldsymbol{H}$, and $\boldsymbol{U}_L^T := \boldsymbol{U}_{N,M_L}$ for $\boldsymbol{L}$, and using a hat to denote the rank r bi-fidelity or low-fidelity reconstruction, respectively, of the matrix. Both of these use only the realizations of the model, and neither of these involve computing PCE coefficients. The transposes are to avoid making changes to the theorem, and do not effect the outcome as the $\|A\|_2 = \|A^T\|_2$. I've also changed $\rho_k(\tau)$ as as to avoid needing to put minimums in the later estimates.

Theorem 0.0.1. *For any $\tau \geq 0$, let*

$$\epsilon(\tau) = \|\boldsymbol{U}_H^T\boldsymbol{U}_H - \tau\boldsymbol{U}_L^T\boldsymbol{U}_L\|_2. \tag{9}$$

Let $\bar{\boldsymbol{U}}_H$ and $\bar{\boldsymbol{U}}_L$ be corresponding static coefficient bi-fidelity estimates of rank r with coefficients $\boldsymbol{C}_L$, and let σ_k denote the kth largest singular value of $\boldsymbol{U}_L$. Then,

$$\|\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H\|_2 \leq \rho_k(\tau); \tag{10}$$

where $\rho_k(\tau)$ is defined by

$$\rho_k(\tau) := \min_{\tau, k \leq rank(\boldsymbol{U}_L)} (1 + \|\boldsymbol{C}_L\|)\sqrt{\tau\sigma_{k+1}^2 + \epsilon(\tau)} + \|\boldsymbol{U}_L - \bar{\boldsymbol{U}}_L\|\sqrt{\tau + \epsilon(\tau)\sigma_k^{-2}}. \tag{11}$$

When $k = rank(\boldsymbol{A}_L)$, we set $\sigma_{k+1} = 0$.

Remark 0.0.1. *I note again that this theorem does not easily apply to this algorithm, as the coefficient matrix $\boldsymbol{C}_L$ is not reused for this bound. However, this bound can still provide an estimate of truncation error. This estimate relies on the Gramian estimate, which relies on high-fidelity samples, and it may be more efficient to simply perform the a posteriori error analysis with the new $\boldsymbol{C}$.*

We also have the theorem from our ℓ_2 paper, that I find more easily applied, though other versions can be used. First, let's define the coherence. As we aren't using importance sampling, this can be simplified by removing the weight function. For presentation, I will ignore the more complicated coherence, and in the paper that complicated coherence has a the probability for $\mathcal{E}$ having an extra $1/r$ term which could be quite large. Let the coherence be defined as

$$\mu_2 := \sup_{\boldsymbol{\xi} \in \Omega} \sum_{k=1}^{r} |\eta_k(\boldsymbol{\xi})|^2. \tag{12}$$

We note that under coherence optimal conditions, $\mu_2 = r$, and such conditions can be guaranteed by importance sampling. Additionally, define the truncation error $\delta_i(\boldsymbol{\xi})$ to be the function achieving the minimum L_2 distance between $u_i(\boldsymbol{\xi})$

and the space of approximations from linear combinations of the basis functions $\{\eta_k(\boldsymbol{\xi})\}_{k=1}^r$. That is δ is the ϵ of the ℓ_2 paper, which has been changed as it may be confused with the $\epsilon(\tau)$ used in the bi-fidelity estimate. We now present Theorem 2.1 of the ℓ_2 paper. We note that the error ν_i in this theorem is the useful bound that we seek.

Theorem 0.0.2. *Let*

$$\hat{u}_i(\boldsymbol{\Xi}) = \sum_{k=1}^{r} \hat{c}(i,k)\eta_k(\boldsymbol{\xi}), \tag{13}$$

where $\hat{c}_i$ is the least-squares solution. It follows that for $\mathcal{E}$, which is independent of i, and is a sampling event that occurs with probability

$$\mathbb{P}(\mathcal{E}) \geq 1 - 2r\,exp(-0.1 N_H \mu_2^{-1}), \tag{14}$$

that

$$\nu_i := \mathbb{E}\left(\|u_i(\boldsymbol{\Xi}) - \hat{u}_i(\boldsymbol{\Xi})\|^2_{L_2(\Omega,f)}; \mathcal{E}\right) \leq \left(1 + \frac{4\mu_2}{N_H}\right)\mathbb{E}(\delta_i^2(\boldsymbol{\Xi})), \tag{15}$$

where μ_2 is as in (12), and

$$\mathbb{E}\left(X; \mathcal{E}\right) = \int_{\mathcal{E}} X(\boldsymbol{\xi}) f(\boldsymbol{\xi}) d\boldsymbol{\xi} = \mathbb{E}(X|\mathcal{E})\mathbb{P}(\mathcal{E}) \tag{16}$$

denotes the expectation restricted to the event (also known as restricted expectation), and is closely related to conditional expectation.

Remark 0.0.2. *I note here that this error bound is for each point in space. The error from the bi-fidelity approximation can be concentrated in certain spatial regions, and this is likely to apply to the PCE approximations for these M_H points as well. The error from each point can be accumulated, as we do in the final corollary here.*

Now we have the new, original, theorem that ties these approximations together.

Theorem 0.0.3. *For $\mathcal{E}$ as in Theorem 0.0.2, and from that theorem, it follows that,*

$$\nu_i := \mathbb{E}\left(\|u_i(\boldsymbol{\Xi}) - \hat{u}_i(\boldsymbol{\Xi})\|^2_{L_2(\Omega,f)}; \mathcal{E}\right) \tag{17}$$

$$\leq \left(1 + \frac{4\mu_2}{N_H}\right)\left(\frac{\Theta_{N_H}}{N_H}\|\boldsymbol{U}_H(:,i) - \hat{\boldsymbol{U}}_H(:,i)\|_2^2\right), \tag{18}$$

Here $\boldsymbol{U}_H$ is the $N_H \times M_H$ matrix of high-fidelity data; $\boldsymbol{C}_L$ is the coefficient matrix associated with $\hat{\boldsymbol{U}}_H$, which is the bi-fidelity approximation to $\boldsymbol{U}_H$; and Θ_{N_H} is a random variable that converges almost surely to 1 as $N_H \to \infty$.

Remark 0.0.3. *We note that ν_i does not tend to zero as $N_H \to \infty$. The first term tends to 1, as μ_2 is finite and does not vary with N_H. However, the second term's $\|\boldsymbol{U}_H(:,i) - \hat{\boldsymbol{U}}_H(:,i)\|_2^2$, grows with N_H, such that the second term converges to a constant, a mean squared error in approximation from using the r basis functions to approximate $\boldsymbol{U}_H$, derived from $\mathbb{E}(\delta_i^2(\boldsymbol{\Xi}))$.*

We also note that Θ_{N_H} depends on the realizations of the random variables $u(\boldsymbol{\Xi})$, and the bi-fidelity reconstruction which fill the vector $\boldsymbol{U}_H(:,i) - \hat{\boldsymbol{U}}_H(:,i)$. Θ_{N_H} is a multiplicative correction to correct the sample mean to the true mean.

Finally, though this bound is ultimately simple, and has a moment estimate in it that is stochastic, it is rather tight. The corollary that uses Theorem 0.0.1, has looser bounds as the bi-fidelity bound is loose, but they are as tight as that loose bound can reasonably permit.

Proof:
We begin with (15), seeking to estimate the truncation error $\mathbb{E}(\delta_i^2(\boldsymbol{\Xi}))$. Recall that, $\boldsymbol{U}_H(k,i)$ contains $u_i(\boldsymbol{\xi}_k)$, and that $\hat{\boldsymbol{U}}_H(k,i)$ contains the corresponding $\hat{u}_i(\boldsymbol{\xi}_k)$ computed via the bi-fidelity PCE. It follows that $\|\boldsymbol{U}_H(:,i) - \hat{\boldsymbol{U}}_H(:,i)\|_2^2$ is an estimate for $\mathbb{E}(\delta_i^2(\boldsymbol{\Xi}))$, where Θ_{N_H} thus converges to 1 almost surely as $N_H \to \infty$ by the strong law of large numbers. ■

We then have the following corollary, where we replace the high-fidelity bound with the associated bi-fidelity bound.

Corollary 1. *Under the conditions of Theorem 0.0.3, it follows that,*

$$\nu_i \leq \left(1 + \frac{4\mu_2}{N_H}\right)\left(\frac{X_{N_H}}{N_H}\rho_k^2(\tau)\right); \tag{19}$$

$$\sum_{i=1}^{M_H} \nu_i \leq \left(1 + \frac{4\mu_2}{N_H}\right)\left(\frac{r\bar{X}_{N_H}}{N_H}\rho_k^2(\tau)\right). \tag{20}$$

Here $\rho_k(\tau)$ is as in Theorem 0.0.1; X_{N_H}, and $\bar{X}_{N_H}$ are random variables which converge a.s. to values in $(0,1]$; r is the rank of $\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H$, and the rest is as in Theorem 0.0.3.

Proof:
Note that for Y_{N_H} a random variable that depends on N_H,

$$\|\boldsymbol{U}_H(:,i) - \hat{\boldsymbol{U}}_H(:,i)\|_2^2 \leq Y_{N_H}\|\boldsymbol{U}_H(:,i) - \bar{\boldsymbol{U}}_H(:,i)\|_2^2, \tag{21}$$

where we recall that the difference between $\hat{\boldsymbol{U}}_H$ and $\bar{\boldsymbol{U}}_H$ is that the former recomputes the coefficient matrix using the high-fidelity data while the latter does not, using the coefficients computed using low-fidelity data. As a result, the LHS converges to $\mathbb{E}(\delta_i^2(\boldsymbol{\Xi}))$, while the RHS converges to an unknown finite quantity. As the LHS converges to the minimum possible value, the random

variable Y_{N_H} converges to some unknown value in $(0, 1]$. Using a a simple matrix norm inequality, and Theorem 0.0.1,

$$\|\boldsymbol{U}_H(:,i) - \bar{\boldsymbol{U}}_H(:,i)\|_2 \leq \|\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H\|_2, \tag{22}$$

$$\leq \rho_k(\tau). \tag{23}$$

In (19), we have $X_{N_H} = \Theta_{N_H} Y_{N_H}$, where Θ_{N_H} is as in Theorem 0.0.3. As Θ_{N_H} converges a.s to 1 and Y_{N_H} converges a.s. to some unknown value in $(0, 1]$, X_{N_H} converges a.s. to the same unknown value as Y_{N_H}.

To show (20), we note that

$$\sum_{i=1}^{M_H} \nu_i \leq \sum_{i=1}^{M_H} \left(1 + \frac{4\mu_2}{N_H}\right) \left(\frac{\Theta_{N_H}(i)}{N_H} \|\boldsymbol{U}_H(:,i) - \hat{\boldsymbol{U}}_H(:,i)\|_2^2\right), \tag{24}$$

where the $\Theta_{N_H}(i)$ are a set of M_H random variables, each convergent to 1 a.s. As a result the maximum of these M_H random variables, which we refer to as $\bar{\Theta}_{N_H}$ also converges to 1 a.s. Having $\bar{X}_{N_H}$ be similarly defined as the maximum of the X_{N_H} above, it follows that $\bar{X}_{N_H}$ converges a.s. to some value in $(0, 1]$. We thus have

$$\sum_{i=1}^{M_H} \nu_i \leq \sum_{i=1}^{M_H} \left(1 + \frac{4\mu_2}{N_H}\right) \left(\frac{\bar{\Theta}_{N_H}}{N_H} \|\boldsymbol{U}_H(:,i) - \hat{\boldsymbol{U}}_H(:,i)\|_2^2\right), \tag{25}$$

$$= \left(1 + \frac{4\mu_2}{N_H}\right) \left(\frac{\bar{\Theta}_{N_H}}{N_H} \sum_{i=1}^{M_H} \|\boldsymbol{U}_H(:,i) - \hat{\boldsymbol{U}}_H(:,i)\|_2^2\right), \tag{26}$$

$$\leq \left(1 + \frac{4\mu_2}{N_H}\right) \left(\frac{\bar{X}_{N_H}}{N_H} \sum_{i=1}^{M_H} \|\boldsymbol{U}_H(:,i) - \bar{\boldsymbol{U}}_H(:,i)\|_2^2\right), \tag{27}$$

$$= \left(1 + \frac{4\mu_2}{N_H}\right) \left(\frac{\bar{X}_{N_H}}{N_H} \|\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H\|_F^2\right), \tag{28}$$

$$\leq \left(1 + \frac{4\mu_2}{N_H}\right) \left(\frac{r\bar{X}_{N_H}}{N_H} \rho_k^2(\tau)\right), \tag{29}$$

where the last line shows (20) and follows from Theorem 0.0.1 and the matrix inequality

$$\|\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H\|_F \leq \sqrt{r} \|\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H\|_2, \tag{30}$$

for r the rank of $\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H$. ■.

One practical enhancement that we can do is to remove the random variables from these bounds, as random variables can be difficult to interpret correctly, especially for practical applciations. The most natural way to proceed seems to be through the Berry-Esseen Theorem. First, define two random variables, to which the Berry-Esseen Bound will be applied.

$$\boldsymbol{V}_{i,j} = |\boldsymbol{U}_H(j,i) - \hat{\boldsymbol{U}}_H(j,i)|^2; \qquad \boldsymbol{W}_j = \sum_{i=1}^{M_H} \boldsymbol{V}_{i,j}. \tag{31}$$

We note that the $\{\boldsymbol{V}_{i,j}\}_{j=1}^{N_H}$ and $\{\boldsymbol{W}_j\}_{j=1}^{N_H}$ are independent and identically distributed for each i.

We also define the moments of $\boldsymbol{V}_{i,j}$, and similarly for $\boldsymbol{W}_j$

$$\mu_{\boldsymbol{V}_i} = \mathbb{E}(\boldsymbol{V}_{i,j}); \qquad \mu_{\boldsymbol{W}} = \mathbb{E}(\boldsymbol{W}_j); \tag{32}$$

$$\sigma^2_{\boldsymbol{V}_i} = \mathbb{E}|\boldsymbol{V}_{i,j} - \mu(\boldsymbol{V}_{i,j})|^2; \qquad \sigma^2_{\boldsymbol{W}} = \mathbb{E}|\boldsymbol{W}_j - \mu(\boldsymbol{W}_j)|^2; \tag{33}$$

$$\rho_{\boldsymbol{V}_i} = \mathbb{E}|\boldsymbol{V}_{i,j} - \mu(\boldsymbol{V}_{i,j})|^3; \qquad \rho_{\boldsymbol{W}} = \mathbb{E}|\boldsymbol{W}_j - \mu(\boldsymbol{W}_j)|^3. \tag{34}$$

We note that as the realizations are identically distributed, there is no difference based on j. We define the appropriately normalized random variables (to mean zero, unit variance) as

$$\tilde{\boldsymbol{V}}_i = N_H^{-1/2} \sum_{j=1}^{N_H} \frac{\boldsymbol{V}_{i,j} - \mu_{\boldsymbol{V}_i}}{\sigma_{\boldsymbol{V}_i}}; \qquad \tilde{\boldsymbol{W}} = N_H^{-1/2} \sum_{j=1}^{N_H} \frac{\boldsymbol{W}_{i,j} - \mu_{\boldsymbol{W}_i}}{\sigma_{\boldsymbol{W}_i}}. \tag{35}$$

We restate the Berry-Esseen Theorem for $\tilde{\boldsymbol{V}}_i$, and $\tilde{\boldsymbol{W}}$.

Theorem 0.0.4. *Let $F_{\tilde{\boldsymbol{V}}_i}(\cdot)$ be the cumulative distribution for $\tilde{\boldsymbol{V}}_i$, and $\Phi(\cdot)$ the cumulative distribution function for the standard normal random variable. There exists a positive constant $C \leq 0.4748$ such that for all t,*

$$|F_{\tilde{\boldsymbol{V}}_i}(t) - \Phi(t)| \leq \frac{C\rho_{\boldsymbol{V}_i}}{\sigma^3_{\boldsymbol{V}_i}\sqrt{N_H}}; \qquad |F_{\tilde{\boldsymbol{W}}}(t) - \Phi(t)| \leq \frac{C\rho_{\boldsymbol{W}}}{\sigma^3_{\boldsymbol{W}_i}\sqrt{N_H}}. \tag{36}$$

Remark 0.0.4. *There are many estimates for the constant C here, I chose only the smallest one that I found on wikipedia. The citation should be one for the original Theorem and one for whichever C is chosen, or multiple if you want to cite a bunch of constants. There are a few highly related Theorems that aren't quite in this form, but similarly enough.*

Corollary 2. *Under the conditions of Theorem 0.0.3, and Corollary 2, it follows that, for any t and with probabilty*

$$p(t) \geq \Phi(t) - \frac{C\rho_{\boldsymbol{V}_i}}{\sigma^3\sqrt{N_H}}, \tag{37}$$

that all of the following bounds hold.

$$\nu_i \leq \left(1 + \frac{4\mu_2}{N_H}\right)\left(\mu_{\boldsymbol{V}_i} + \frac{t\sigma_{\boldsymbol{V}_i}}{\sqrt{N_H}}\right). \tag{38}$$

Similarly, for any t, and with probability

$$p(t) \geq \Phi(t) - \frac{C\rho_{\boldsymbol{W}}}{\sigma^3\sqrt{N_H}}, \tag{39}$$

$$\sum_{i=1}^{M_H} \nu_i \leq \left(1 + \frac{4\mu_2}{N_H}\right)\left(\mu_{\boldsymbol{W}} + \frac{t\sigma_{\boldsymbol{W}}}{\sqrt{N_H}}\right). \tag{40}$$

Remark 0.0.5. *This proof follows directly from algebra on (19) and (19) the Berry-Esseen bound, and the definitions of (35). I omit any details here, but we might add them in the paper. This bound is tight when N_H is large enough N_H that the Berry-Esseen corrected probability $p(t)$ is not very different than $\Phi(t)$.*

However, with this idea N_H is not likely to be vary large, so this may not be very tight, generally. There is no other reasonable distribution to compare to except the normal distribution, at least not without extraordinary problem knowledge. The distribution of $\boldsymbol{V}_i$ and, especially $\boldsymbol{W}_i$ are themselves likely not far from normality. In such a case, the corresponding ρ will be small. Estimating ρ can be a done via a sample average from the definition of (34). The estimates of the mean and the variance are also reasonable replacements for the theoretical values here.

Felix Newberry, October, 2020: Bi-fidelity PCE - Numerical Testing

Start: numerical results for practical extension

Updated with numerical results for practical enhancement. Set $C = 0.4748$ and computed the probabilities and bounds of equations (37) to (40) for each numerical example. For $t = 0.95$ the probability was 0.7228 for all four examples. Table 1 has been updated with the efficacy of equation (40). Likewise, Figures 1, 6, 11 and 16 have been updated with the computed bound from equation (38).

Practical enhancement observations:

- Generally the practical bound performs very well, with two exceptions
- In the airfoil the practical bound underestimates the true error
- For the gas turbine vertical line QoI, both the former bound and the practical bound fail for some values at high x
- The unexpected behavior in both the airfoil and gas turbine is resolved if either the number of samples is increased, or we take the average over 30 repetitions. Figures 21, 22, 23 and 24 deomonstrate this.
- In general, behavior seems consistent across changes in N_H and r, so I think we're justified to show just one case (repeated 30 times)

Intention is to present equation (29) as the analytical bound that we can look to for intuition about how the error behaves. In the reulst section we show via efficacy and plots of Y and Θ that this bound does not make sense to apply in practice. We instead apply the practical enhancement. Proposed content to include in paper, where all numerical results are average over 30 repetitions:

- All the theory
- Table 2 - with Eqn 40 added, and the content distributed to each numerical example
- Table 3 - with the content distributed to each numerical example
- For each numerical example, plots of Y, $Theta$ and the error bound per point. Undecided whether to include 'Bound 27' in the bound plot. For the LDC it doesn't look great

End: numerical results for practical extension

Here's a quick write up of the numerical results to date using Jerrad's error bound. The results are very encouraging. I've tested the bound in its present form on the three numerical examples from the paper, namely the lid driven cavity, gas turbine and NACA airfoil. In each example I've compared terms from the inequalities in equations (26) through (29).

The primary takeaway is that if we apply equation (27) as a bound on each point, then the bound is very tight. Once we 1) take the maximum values of Θ_{N_H} and Y_{N_H} to compute $\bar{X}_{N_H}$, and 2) introduce $\|\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H\|_F^2 \leq r\rho_k^2(\tau)$ the bound loosens considerably, in the case of the lid-driven cavity it loosens drastically. Note that in this study when I compute equation(27) I use the optimal Θ_{N_H} and Y_{N_H} values for each point in M_H.

An important caveat is that to date I've tested the bound for only one (N_H, r) pair and only one set of N_H samples. Testing prior to publication will certainly include many repetitions of N_H samples, and look at varying (N_H, r).

Table 1 summarizes the change in efficacy for each test case from equation (27) to (29). We observe that the lid driven cavity and the gas turbine vertical line see a very large disparity in efficacy from the two equations. Gas turbine cylinder surface and the airfoil, exhibit comparatively better efficacies in equation (29), but still show a notable gap from equation (27).

In Table 3 we examine to what extent the disparities of Table 1 can be accounted for by the bounding the matrix ID bi-fidelity estimate by $r\rho_k(\tau)^2$. In general, it appears that this inequality is responsible for approximately one order of magnitude of the final efficacy. We infer that the remaining disparity is due to taking the maximum values of Θ_{N_H} and Y_{N_H} to compute $\bar{X}_{N_H}$.

	Eqn (29)	Eqn (27)	Eqn (40)
LDC	207	2.00	2.27
Gas Turbine Vertical Line	107	1.58	1.56
Gas Turbine Cylinder	36.2	2.00	2.05
Airfoil	23.9	1.35	0.646

Table 1: Comparison of error bound efficacy calculated from the ratio of the square root of the RHS of equations (27),(29) and (40) to the square root of the LHS. The bound in equation (27) was computed with X_{N_H} as opposed to taking the maximum of Θ_{N_H} and Y_{N_H} to find $\bar{X}_{N_H}$.

	Eqn (29)	Eqn (27)	Eqn (40)
LDC	247	1.87	1.38
Gas Turbine Vertical Line	132	2.04	1.75
Gas Turbine Cylinder	32.9	2.30	1.67
Airfoil	20.0	1.97	1.1

Table 2: Comparison of error bound efficacy calculated from the ratio of the square root of the RHS of equations (27),(29) and (40) to the square root of the LHS. The bound in equation (27) was computed with X_{N_H} as opposed to taking the maximum of Θ_{N_H} and Y_{N_H} to find $\bar{X}_{N_H}$. Results were calculated as the average of 30 repetitions.

	$\sqrt{r\rho_k^2(\tau)}$	$\|\boldsymbol{U}_H - \bar{\boldsymbol{U}}_H\|_F$
LDC	4.44e-4	1.61e-4
Gas Turbine Vertical Line	1.49e-1	2.06e-2
Gas Turbine Cylinder	3.65e-1	3.09e-2
Airfoil	6.14e-2	8.64e-3

Table 3: Comparison of inequality between equations (28) and (29). To facilitate relating this table to the efficacy in Table 1, the square root has been taken.

In the following I've given a brief description of each numerical test and figures showing the performance of equation (27) and values in Θ_{N_H} and Y_{N_H}. For complete details of each model please refer to the paper draft.

In each test case we calculate Θ_{N_H} from equation (18) by computing $\left(\frac{\Theta_{N_H}}{N_H}\|\boldsymbol{U}_H(:,i) - \hat{\boldsymbol{U}}_H(:,i)\|_2^2\right)$ for a limited N_H and $N_{H\infty}$, the maximum available number of samples as a surrogate for $N_H \to \infty$. In practice, we will need some estimate of Θ_{N_H}. Θ_{N_H} is a vector of length M, with an entry for each spatial coordinate.

We can calculate Y_{N_H} directly from equation (21). This is the ratio of the PC bi-fidelity error to the matrix ID error squared. Like Θ_{N_H}, Y_{N_H} is a vector of length M. To compute $\bar{X}_{N_H}$ we take the maximum of both Θ_{N_H} and Y_{N_H}.

General observations from figures:

- The assumption that matrix ID is a looser estimate than the PC bi-fidelity, and hence that Y_{N_H} is in $(0, 1]$ does not hold for the LDC, but holds well for the gas turbine and the airfoil.
- All three numerical test cases give reasonable values for Θ_{N_H}
- LDC Y_{N_H} ratio maximum is responsible for a large part of the bound - ratio behaves poorly because matrix ID estimate has very small error for some points
- Gas turbine vertical line shows a very large difference between the maximum Y_{N_H} and most of the Y_{N_H} values likely makes up a large part of the final bound.
- The major part of the bounds looseness is accounted by bounding above with matrix ID, first in the introduction of Y_{N_H} and later using $\rho_k(\tau)$. I'm not certain whether this would be mathematically sound, but stopping at equation (26) would yield much better efficacy.
- Estimating Θ_{N_H} using only a limited number of samples will be necessary to employ the bound in practice.

Numerical Test Case I: Lid Driven Cavity (LDC):

In this test case the LDC QoI is vertical velocity at $M = 65$ points through the cavity center, $y = 0.5$ from $x = 0$ to $x = 1$. We estimate $N = 200$ samples with $N_H = 20$ high-fidelity samples, a bi-fidelity rank of $r = 3$, $R = 13$ samples to compute the matrix ID error bound $\rho_k(\tau)$ and $N_{H\infty} = 2200$ to estimate Θ_{N_H}.

We examine the origin of the loose efficacy with Figure 1 which plots the mean error for each point alongside the corresponding point specific bound from equation (27). Figure 2 shows the corresponding efficacy of equation (27) for each point. We observe that prior to the summation and introduction of Y_{N_H} the bound is quite tight, with an efficacy close to 2 as opposed to 200 (refer back to Table 1). Evidently a combination of computing the maximum value of Θ_{N_H} and Y_{N_H}, and the introduction of $\rho_k(\tau)$ must account for this disparity.

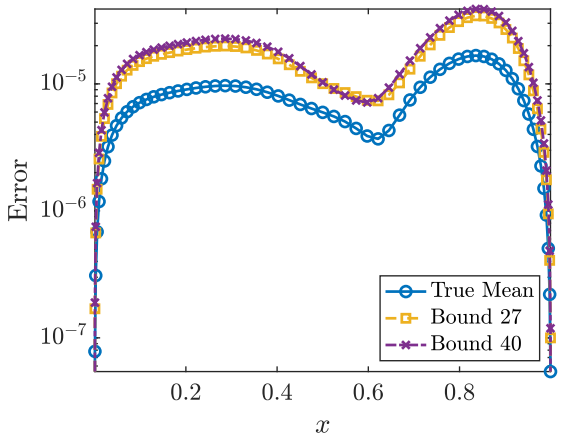

Figure 1: LDC Comparison of the true mean error and the bound calculated from equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken.

Next we plot Θ_{N_H} as a function of x in Figure 3. We find that Θ exhibits entirely reasonable values, ie close to 1, and the maximum value is no cause for concern.

Finally, we plot Y_{N_H} as a function of x in Figure 4. It is immediately apparent that several very large peaks occur in Y, while most of the values are much more reasonable. To examine this we plot the two terms that comprise Y_{N_H} from equation (21) in Figure 5. We find that the error of the matrix ID estimate in Figure 5 tends to be smaller than the error of the PC bi-fidelity estimate. It is clear that where the matrix ID error is close to 0 Y_{N_H} becomes very large and in turn causes the bound summed over all points that uses the maximum of Y to perform poorly in efficacy.

Numerical Test Case II: Gas Turbine:

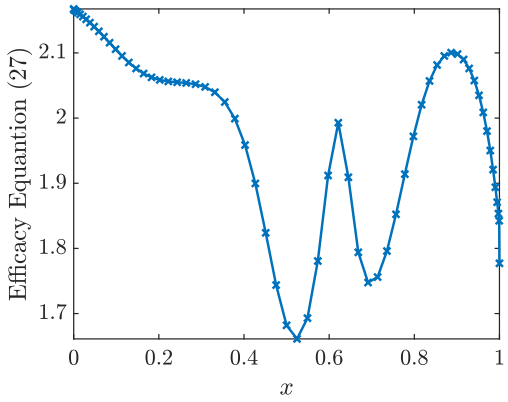

Figure 2: LDC Error bound efficacy in equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken.

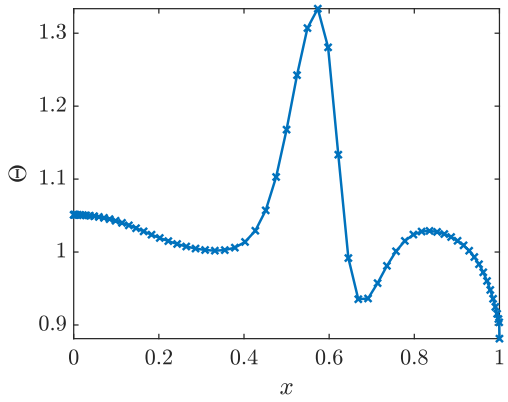

Figure 3: LDC Θ_{N_h} as a function of location in x.

For full details of the Gas Turbine model please refer to the paper draft. The problem can be thought of as heated flow past a cylidner. In this test case we consider two separate QoI, the temperature along a vertical line at $x = 0.2$ and temperature around the entire cylinder surface from $-\pi$ to π radians. The vertical line has $M = 128$ points while the cylinder surface has $M = 202$. Both sets of coordinates are represented by x in this study. For both QoI we estimate $N = 100$ samples with $N_H = 20$ high-fidelity samples, a bi-fidelity rank of $r = 8$, $R = 18$ samples to compute the matrix ID error bound $\rho_k(\tau)$ and $N_{H_\infty} = 200$.

Gas turbine temperature along vertical line figure summary:

- Figure 6 plots mean error for each point alongside the corresponding point

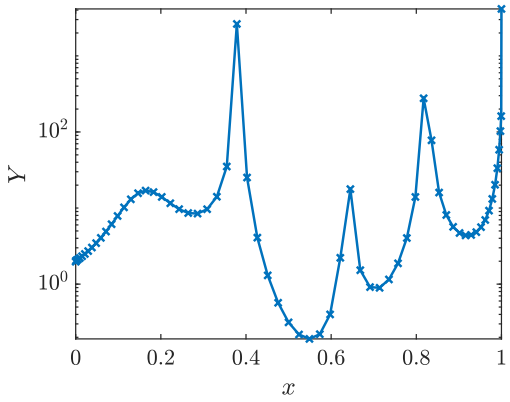

Figure 4: LDC Y_{N_h} as a function of location in x.

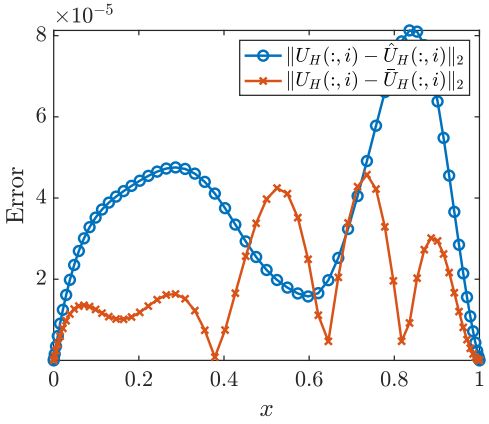

Figure 5: LDC Terms that comprise Y_{N_h} as a function of location in x.

specific bound from equation (27)

- Figure 7 shows efficacy of equation (27) for each point
- Figure 8 plots Θ_{N_H} as a function of x
- Figure 9 plots Y_{N_H} as a function of x
- Figure 10 plots the two terms that comprise Y_{N_H} from equation (21)

Gas turbine temperature around cylinder surface figure summary:

- Figure 11 plots mean error for each point alongside the corresponding point specific bound from equation (27)

- Figure 12 shows efficacy of equation (27) for each point
- Figure 13 plots Θ_{N_H} as a function of x
- Figure 14 plots Y_{N_H} as a function of x
- Figure 15 plots the two terms that comprise Y_{N_H} from equation (21)

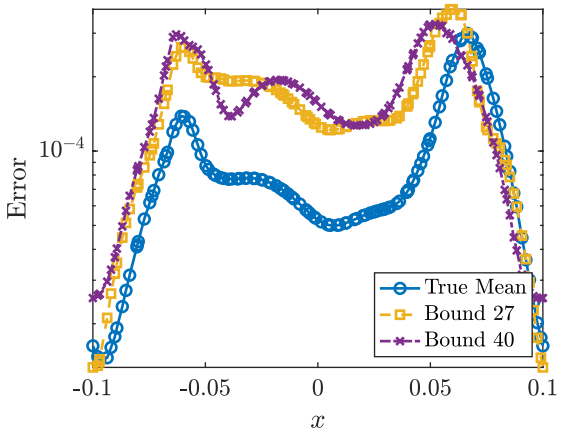

Figure 6: Gas turbine vertical line. Comparison of the true mean error and the bound calculated from equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken.

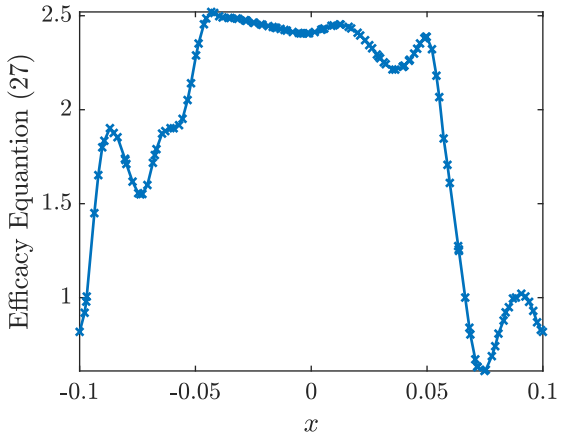

Figure 7: Gas turbine vertical line. Error bound efficacy in equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken.

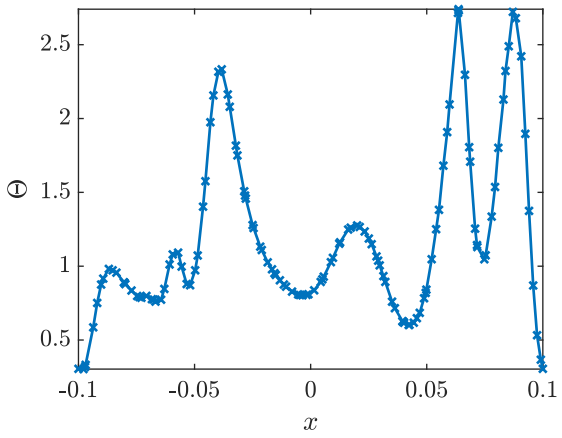

Figure 8: Gas turbine vertical line. Θ_{N_h} as a function of location in x.

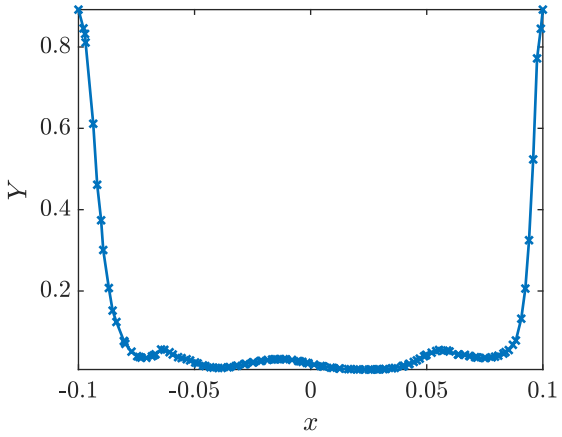

Figure 9: Gas turbine vertical line. Y_{N_h} as a function of location in x.

Numerical Test Case III: Airfoil:

For full details of the Airfoil model please refer to the paper draft. In this test case the QoI is the coefficient of pressure at $M = 200$ points evenly spaced around the airfoil's surface, mapped to x -1 to 1. We estimate $N = 200$ samples with $N_H = 20$ high-fidelity samples, a bi-fidelity rank of $r = 5$, $R = 15$ samples to compute the matrix ID error bound $\rho_k(\tau)$ and $N_{H_\infty} = 500$.

Airfoil figure summary:

- Figure 16 plots mean error for each point alongside the corresponding point specific bound from equation (27)
- Figure 17 shows efficacy of equation (27) for each point

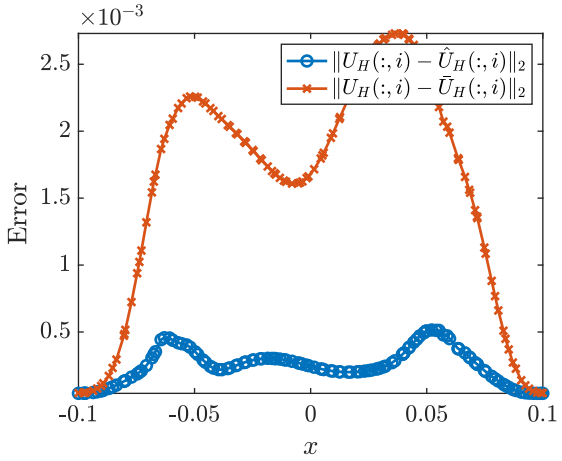

Figure 10: Gas turbine vertical line. Terms that comprise Y_{N_h} as a function of location in x.

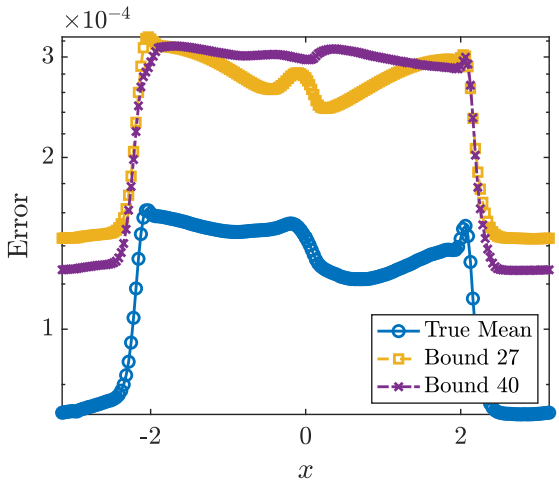

Figure 11: Gas turbine cylinder. Comparison of the true mean error and the bound calculated from equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken.

- Figure 18 plots Θ_{N_H} as a function of x
- Figure 19 plots Y_{N_H} as a function of x
- Figure 20 plots the two terms that comprise Y_{N_H} from equation (21)

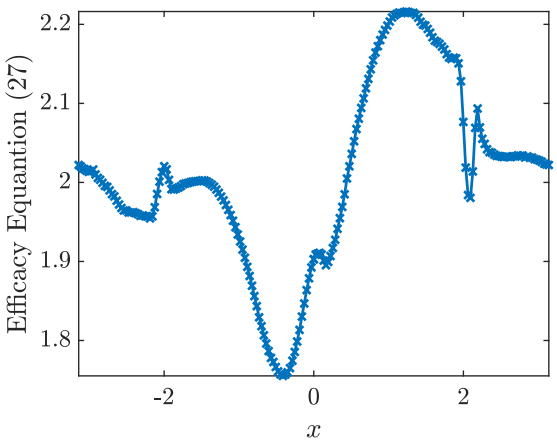

Figure 12: Gas turbine cylinder. Error bound efficacy in equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken.

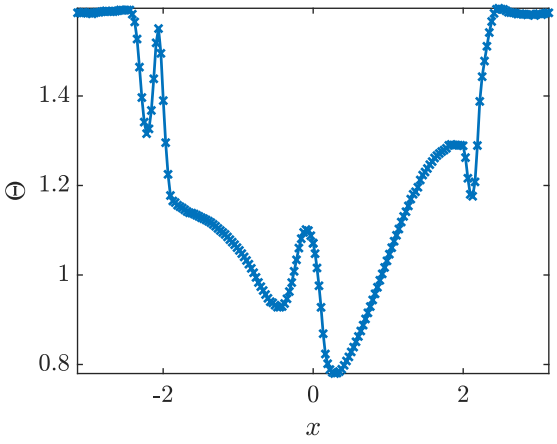

Figure 13: Gas turbine cylinder. Θ_{N_h} as a function of location in x.

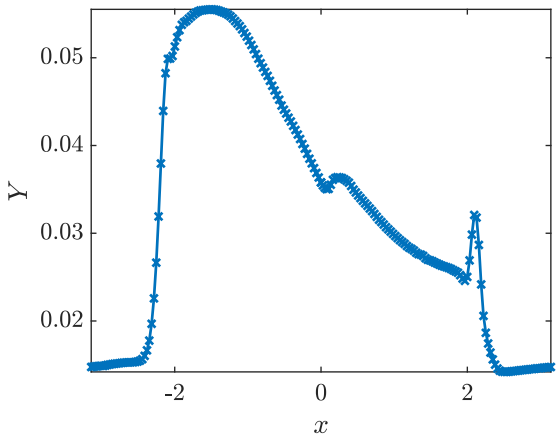

Figure 14: Gas turbine cylinder. Y_{N_h} as a function of location in x.

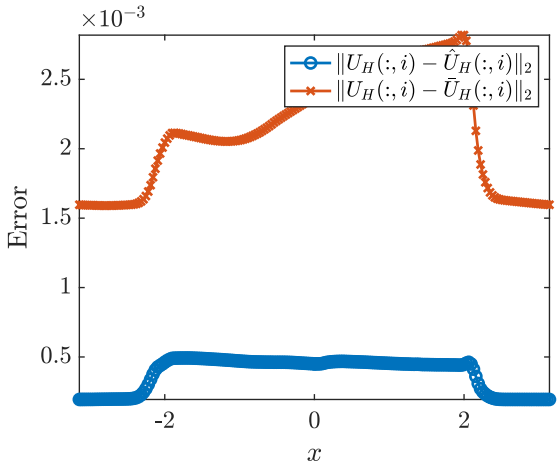

Figure 15: Gas turbine cylinder. Terms that comprise Y_{N_h} as a function of location in x.

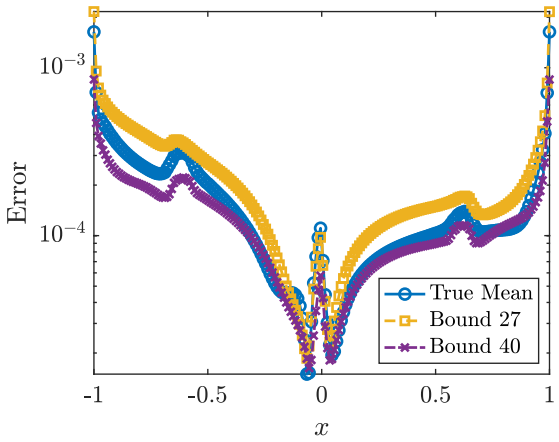

Figure 16: Airfoil Comparison of the true mean error and the bound calculated from equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken.

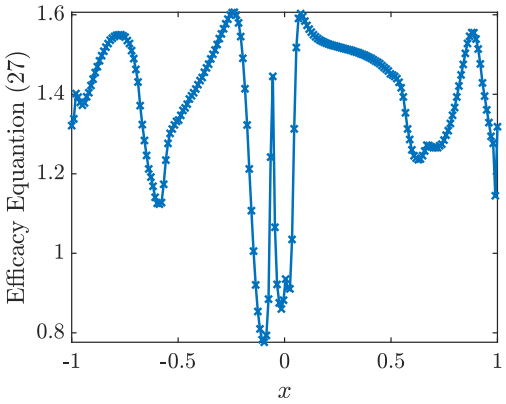

Figure 17: Airfoil Error bound efficacy in equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken.

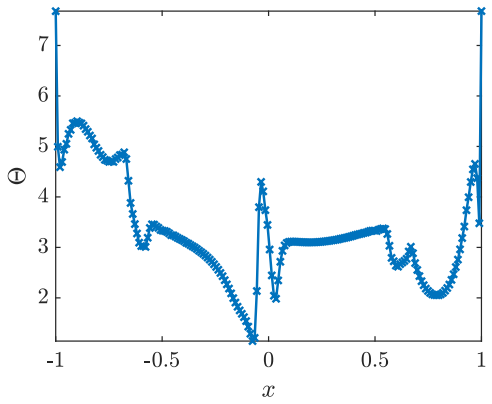

Figure 18: Airfoil Θ_{N_h} as a function of location in x.

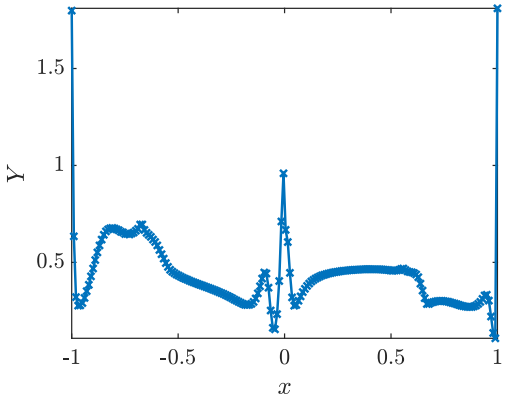

Figure 19: Airfoil Y_{N_h} as a function of location in x.

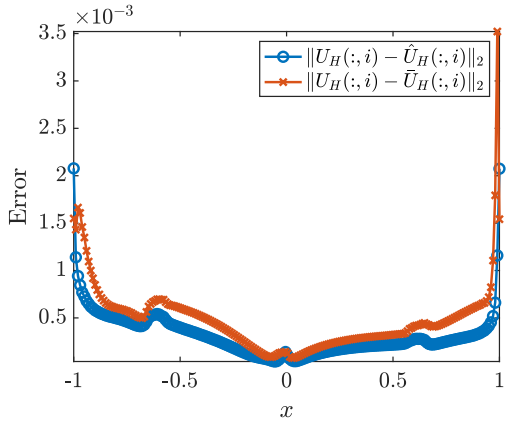

Figure 20: Airfoil Terms that comprise Y_{N_h} as a function of location in x.

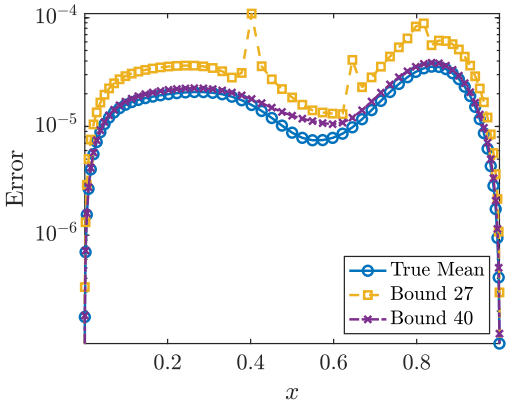

Figure 21: LDC Comparison of the true mean error and the bound calculated from equations (27) and (38) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken. The error bound is computed from the average of 30 repetitions.

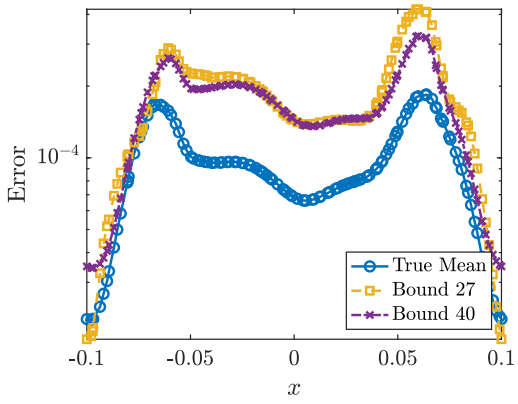

Figure 22: Gas turbine vertical line. Comparison of the true mean error and the bound calculated from equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken.

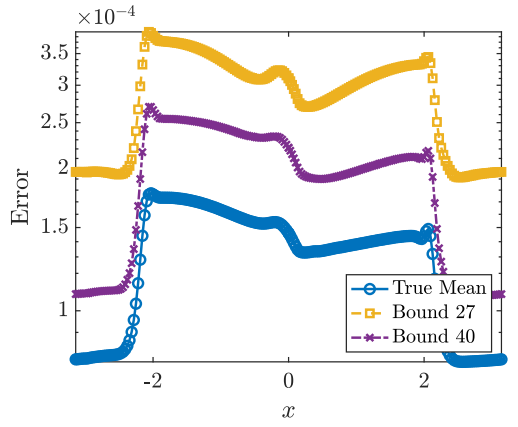

Figure 23: Gas turbine cylinder. Comparison of the true mean error and the bound calculated from equation (27) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken. The error bound is computed from the average of 30 repetitions.

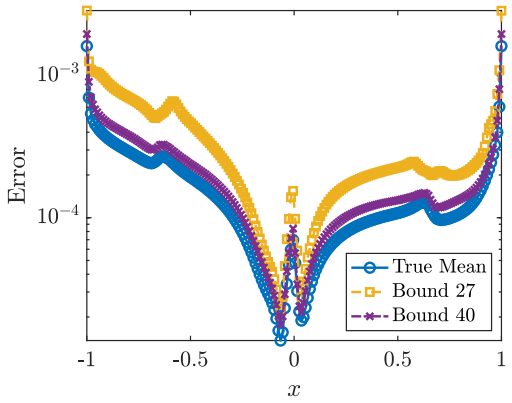

Figure 24: Airfoil Comparison of the true mean error and the bound calculated from equations (27) and (38) for all M points. The mean error for each point is computed from $N = 200$ samples and the square root has been taken. The error bound is computed from the average of 30 repetitions.