Meta variance reduction for Monte Carlo estimation of energetic particle confinement during stellarator optimization

Frederick Law[a], Antoine Cerfon[a], Benjamin Peherstorfer[a], Florian Wechsung[a]

[a] *Courant Institute of Mathematical Sciences, New York University, United States of America*

Abstract

This work introduces meta estimators that combine multiple multi-fidelity techniques based on control variates, importance sampling, and information reuse to yield a quasi-multiplicative amount of variance reduction. The proposed meta estimators are particularly efficient within outer-loop applications, such as reliability-based design and shape optimization, when the input distribution of the uncertainties changes with the outer-loop iteration. We derive asymptotic bounds of the variance reduction of the meta estimators in the limit of convergence of the outer-loop applications. We demonstrate the meta estimators, using data-driven surrogates and biasing densities, on a design problem under uncertainty from nuclear fusion, namely the optimization of stellarator coil designs to maximize the estimated confinement of energetic particles. The meta estimators outperform all of their constituent variance reduction techniques alone, ultimately yielding two orders of magnitude speedup compared to standard Monte Carlo estimation at the same computational budget.

Keywords: multifidelity, importance sampling, information reuse, variance reduction, stellarator

1. Introduction

Stellarators are a promising type of magnetic confinement fusion reactors, which address several of the challenges facing tokamaks, the other strong contenders for commercially viable magnetic confinement fusion energy [1, 2, 3]. Stellarators are complex machines from an engineering point of view, and are designed via lengthy reactor optimization studies based on computationally expensive multi-physics codes [4, 5, 6, 7, 8]. One of the key considerations for these design studies is the confinement of energetic alpha particles, born from the fusion of the deuterium and tritium nuclei in the reactor. Good confinement is an essential feature to maximize the net self-heating power [9, 10], and therefore obtain favorable power balance [10, 11]. It is also critical to minimize the impact of energetic particle losses on the plasma facing component [1, 12, 9, 13, 11].

The birth of alpha particles is most accurately described as a random process, where the location of birth of a given particle, and the direction of its initial velocity are random variables. Consequently, alpha particle confinement studies are often done following a Monte Carlo approach [14, 15, 16, 17, 18, 19, 20, 21]: a large ensemble of initial conditions for alpha particles is generated by sampling the physically appropriate distributions for birth location and velocity direction, and alpha particles trajectories are computed from these initial conditions; one then estimates the alpha particle confinement statistics of interest via the corresponding Monte Carlo estimates obtained from the trajectories. Since particle trajectories are expensive to compute for the desired level of accuracy, a direct Monte Carlo approach typically is computationally intractable.

Deterministic measures of the quality of alpha particle confinement have been proposed in the recent past [22, 23, 24, 13] to address this limitation. They are less computationally expensive to estimate, and are therefore more practical to include in multi-physics optimization codes. Furthermore, they have proven to be fairly reliable in the regime for which they were designed [24, 13]. Nevertheless, their predictive capability is not perfect, and is limited to specific particle loss channels [24, 13]. In that context, another strategy to achieve strong alpha particle confinement in stellarators is to obtain it as a natural by-product of another highly desirable property of the magnetic configuration targeted during optimization, called quasi-symmetry [25], which also guarantees the confinement of the thermal deuterium-tritium fuel. It was indeed recently shown that magnetic fields with an unprecedented level of quasi-symmetry confine energetic alpha particles

extremely well [18, 19, 20, 26, 21]. Based on these promising results, one could be tempted to conclude that energetic particle confinement codes need not be included in multi-physics stellarator design studies, and that good confinement should simply be verified numerically, via a single expensive Monte Carlo estimation, once an optimized configuration with good quasi-symmetry has been computed. There are two caveats to such a strategy. First, not all optimized stellarators have quasi-symmetric magnetic fields [15, 25, 27, 28, 29], and unlike quasi-symmetric fields, it has not yet been numerically shown that excellent energetic particle confinement naturally follows from the construction of these other types of optimized magnetic fields [29] . Second, the excellent confinement results found recently for quasi-symmetric configurations were obtained for designs that did not account for several engineering constraints and criteria, such as the geometry and location of plasma facing components [30], and blanket design for thermal and neutral shielding as well as tritium breeding and heat exchange [31, 32]. The level of quasi-symmetry of reactor designs accounting for these engineering constraints is likely to be lower than in the physics-driven designs recently published, as the optimization becomes more complex, and physics targets are no longer the only driving objectives. Since alpha particle confinement can degrade rapidly with increasing deviations from quasisymmetry [24], ensuring good confinement performance necessitates the inclusion of reliable measures of confinement in the optimization objectives. The purpose of this work is to present a framework that can accelerate the Monte Carlo estimation of energetic particle confinement, so that its computational cost becomes comparable to other codes in multi-physics, multi-engineering stellarator optimization codes, and so that alpha particle confinement estimation can be integrated in the optimization process.

Specifically, in this article we introduce meta estimators which simultaneously leverage multiple techniques of multifidelity variance reduction to accelerate Monte Carlo estimation. Our meta estimators are based on constituent multifidelity estimators which each yield variance reduction by taking advantage of different aspects of the estimation problem. To reduce variance through correlated model outputs, we use multifidelity Monte Carlo (MFMC) with a data-driven surrogate [33, 34, 35, 36, 37, 38]; see also multi-level Monte Carlo methods [39, 40, 41]. Variance reduction based on control variates has been used extensively in, e.g., kinetic models such as Boltzmann's equation in [42, 43]. Stochastic collocation techniques based on sparse grids and dimension-adaptive surrogate models for benchmark scenarios of plasma micro-turbulence simulations are introduced in [44, 45]. Other estimation techniques based on polynomial chaos and quasi-Monte Carlo methods are investigated in the context of plasma fusion simulations in [46, 47, 48]. We go a step further than using variance reduction based on control variates alone and also use importance sampling [49, 50]. In particular, we use biasing densities that are fitted to data such that they place mass in regions of the input space which we are interested in, akin to multi-fidelity importance sampling techniques [51, 52, 53]. Finally, we include the concept of information reuse [33, 54] that uses estimators of past optimization iterates as control variates at the current iteration. A high-level illustration of this combination process is given in Figure 1.

This manuscript is structured as follows. In Section 2 we review the three existing methods we seek to combine, namely: multifidelity Monte Carlo, importance sampling, and information reuse. One of our contributions is in Section 3 where we adapt the original information reuse and multifidelity information reuse estimators to remain unbiased under changing input distributions, as well as derive asymptotic bounds for their variances. Our primary contribution is in Section 4, where we combine three multi-fidelity estimators to construct meta estimators. For meta estimators which have our adaptive information reuse estimator as a constituent method, we also derive the related asymptotic variance reduction. In Section 5 we detail the uncertainty propagation problem of energetic alpha particle confinement in stellarators. In Section 6 we present our numerical results using meta estimators on a single NCSX-like configuration [55] and on the optimization trajectory for a new quasi-axisymmetric configuration by Landreman and Paul [18].

2. Preliminaries

We review the standard Monte Carlo, multifidelity Monte Carlo, importance sampling, and information reuse methods.

2.1. Standard Monte Carlo estimation

We are interested in outer-loop applications under input uncertainty, such as design optimization under uncertainty and robust control. We denote the deterministic outer-loop variable as $\boldsymbol{\lambda}^{(k)} \in \mathbb{R}^{N_{\text{loop}}}, N_{\text{loop}} \in \mathbb{N}$,

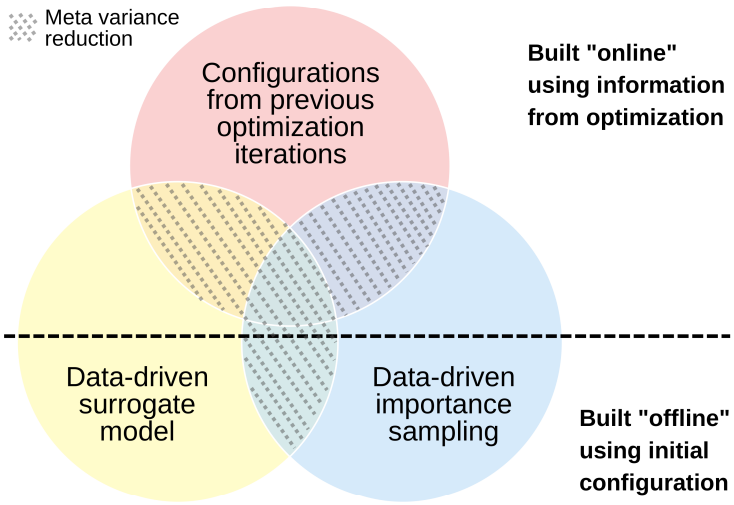

Figure 1: Venn diagram demonstrating the three different estimators to be combined.

which depends on the outer-loop iteration $k = 0, 1, 2 \ldots$ and represents, e.g., the design variable and the control parameters. Because we consider outer-loop applications under uncertainty, we also have a domain of input uncertainty that we denote as $D \subset \mathbb{R}^d$, which is a simply connected, bounded set. Let then $\pi^{(k)}_{\boldsymbol{\lambda}^{(k)}}$ be the probability density function of a probability distribution on D, and let $F^{(k)}_{\boldsymbol{\lambda}^{(k)}} : D \to \mathbb{R}$ be a function that represents a high-fidelity model (HFM) of the system of interest. The density $\pi^{(k)}_{\boldsymbol{\lambda}^{(k)}}$ and the high-fidelity model $F^{(k)}_{\boldsymbol{\lambda}^{(k)}}$ depend on the outer-loop variable $\boldsymbol{\lambda}^{(k)}$. However, for ease of exposition, we will suppress in the notation the explicit dependence of $\pi^{(k)}_{\boldsymbol{\lambda}^{(k)}}$ and $F^{(k)}_{\boldsymbol{\lambda}^{(k)}}$ on $\boldsymbol{\lambda}^{(k)}$ and simply write $\pi^{(k)}$ and $F^{(k)}$ instead. We will subsequently refer to the kth instance of the HFM using outer-loop variables $\boldsymbol{\lambda}^{(k)}$ as the kth HFM. Our goal is estimating $\mathbb{E}_{\pi^{(k)}}[F^{(k)}(\xi)]$ at the outer-loop iterations $k = 0, 1, 2, \ldots$, where $\xi \sim \pi^{(k)}$ is a random variable on D. The computational cost of one evaluation of $F^{(k)}$ is constant for all k and the normalized $\pi^{(k)}$ sampled and evaluated for each k. Assume that $\boldsymbol{\lambda}_k \to \boldsymbol{\lambda}$ as $k \to \infty$.

To estimate $\mathbb{E}_{\pi^{(k)}}[F^{(k)}(\xi)]$ with standard Monte Carlo, consider a computational budget of p HFM evaluations at the kth outer-loop iteration with samples $\xi_1, \ldots, \xi_p$ independent and identically distributed (i.i.d.) from $\pi^{(k)}$. The regular Monte Carlo estimator and its variance are

$$\widehat{F}^{(k)}_{\mathrm{MC},p} = \frac{1}{p}\sum_{i=1}^{p} F^{(k)}(\xi_i), \qquad \mathrm{Var}_{\pi^{(k)}}\left[\widehat{F}^{(k)}_{\mathrm{MC},p}\right] = \frac{\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]}{p}, \tag{1}$$

respectively. The MC estimator is unbiased and so the mean-square error (MSE) is the variance of the estimator. If the variance of $F^{(k)}$ is large, then the number of HFM evaluations p will need to be large to reduce the MSE. But if HFM evaluations are expensive, utilizing large p may be too computationally prohibitive. This is true for each outer-loop iteration k, and therefore using just the MC estimator alone may be intractable for estimating $\mathbb{E}^{(k)}_{\pi}[F^{(k)}]$ during outer-loop applications.

2.2. Multifidelity Monte Carlo

In addition to the HFM $F^{(k)}$, we now also have given a surrogate model $G^{(k)} : D \to \mathbb{R}$. The multifidelity Monte Carlo (MF) estimator [40, 34, 35] leverages $G^{(k)}(\xi)$ as a control variate for estimating the expected value of the HFM $F^{(k)}$. Let $\rho_{\pi^{(k)}}(F^{(k)}, G^{(k)})$ be the Pearson's correlation coefficient, and let w denote the ratio of cost for evaluating $F^{(k)}$ to the cost for evaluating $G^{(k)}$. Then the MF estimator [34] with computational budget equivalent to p HFM evaluations is

$$\widehat{F}^{(k)}_{\mathrm{MF},p} := \widehat{F}_{\mathrm{MF},F^{(k)},G^{(k)},\pi^{(k)},p} = \left(\frac{1}{n}\sum_{i=1}^{n} F^{(k)}(\xi_i)\right) + \alpha\left(\frac{1}{m}\sum_{i=1}^{m} G^{(k)}(\xi_i) - \frac{1}{n}\sum_{i=1}^{n} G^{(k)}(\xi_i)\right) \tag{2}$$

where ξ_i are drawn i.i.d. from $\pi^{(k)}$, n is number of HFM evaluations, m is the number of surrogate model evaluations, and

$$\alpha = \rho_{\pi^{(k)}}(F^{(k)}, G^{(k)}) \sqrt{\frac{\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]}{\mathrm{Var}_{\pi^{(k)}}[G^{(k)}]}}, \quad p = n + \frac{m}{w}, \quad m = n \sqrt{\frac{w \rho_{\pi^{(k)}}(F^{(k)}, G^{(k)})^2}{1 - \rho_{\pi^{(k)}}(F^{(k)}, G^{(k)})^2}}.$$

The variance of the MF estimator is

$$\mathrm{Var}_{\pi^{(k)}}\left[\widehat{F}^{(k)}_{\mathrm{MF},p}\right] = c^{(k)}_{\mathrm{MF}} \frac{\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]}{p}, \quad c^{(k)}_{\mathrm{MF}} = \left(\sqrt{1 - \rho_{\pi^{(k)}}(F^{(k)}, G^{(k)})^2} + \sqrt{\frac{\rho_{\pi^{(k)}}(F^{(k)}, G^{(k)})^2}{w}} \right)^2. \tag{3}$$

A high variance reduction is achieved if the surrogate model $G^{(k)}$ is cheap to evaluate and its output random variable $G^{(k)}(\xi)$ is highly correlated to the HFM output $F^{(k)}(\xi)$. Note that the MF estimator shrinks the variance by taking advantage of correlation between model *outputs*. A key feature to note is that once a surrogate $G^{(k)}$ is selected, the variance reduction $c^{(k)}_{\mathrm{MF}}$ is independent of p.

For the rest of this work "the MF estimator" will refer specifically to $\widehat{F}^{(k)}_{\mathrm{MF},p}$, whereas the more general notation for $\widehat{F}_{\mathrm{MF},F,G,\pi,p}$ will be useful for different choices of F, G, π and p, with different values for α, n, m, and w respectively.

2.3. Importance sampling

While the MF estimator provides variance reduction by taking advantage of correlated model outputs, we now review importance sampling that judiciously puts more weight in regions of the model *input* domain with a biasing distribution to reduce the variance.

Let $\tilde{\pi}^{(k)}$ be the density of a biasing distribution on D that satisfies $\mathrm{supp}(\pi^{(k)}) \subseteq \mathrm{supp}(\tilde{\pi}^{(k)})$. Let further $\widetilde{F}^{(k)} : D \to \mathbb{R}$, defined by $\xi \mapsto F^{(k)}(\xi)\pi^{(k)}(\xi)/\tilde{\pi}^{(k)}(\xi)$ denote the importance weighted HFM. In the following, the cost of evaluating $F^{(k)}(\xi)$ and $\widetilde{F}^{(k)}(\xi)$ is the same for a given ξ, because the cost of evaluating $\pi^{(k)}$ and $\tilde{\pi}^{(k)}$ is typically negligible compared to $F^{(k)}$.

The importance sampling (IS) estimator with a computational budget of p HFM evaluations samples $\xi_1, \ldots, \xi_p$ i.i.d. from $\tilde{\pi}^{(k)}$, evaluates $\widetilde{F}^{(k)}$ for each sample, and takes the sample average:

$$\widehat{F}^{(k)}_{\mathrm{IS},p} = \frac{1}{p} \sum_{i=1}^{p} \widetilde{F}^{(k)}(\xi_i). \tag{4}$$

The variance of the IS estimator is

$$\mathrm{Var}_{\tilde{\pi}^{(k)}}\left[\widehat{F}^{(k)}_{\mathrm{IS},p}\right] = c^{(k)}_{\mathrm{IS}} \frac{\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]}{p}, \qquad c^{(k)}_{\mathrm{IS}} := \frac{\mathrm{Var}_{\tilde{\pi}^{(k)}}[\widetilde{F}^{(k)}]}{\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]}. \tag{5}$$

The IS estimator is unbiased, as $\mathbb{E}_{\tilde{\pi}^{(k)}}[\widetilde{F}^{(k)}] = \mathbb{E}_{\pi^{(k)}}[F^{(k)}]$ provided $\mathrm{supp}(\pi^{(k)}) \subseteq \mathrm{supp}(\tilde{\pi}^{(k)})$. The optimal biasing density is $\tilde{\pi}^{(k)}_* \propto |F^{(k)}|\pi^{(k)}$ in that $\mathrm{Var}_{\tilde{\pi}^{(k)}_*}[F\pi^{(k)}/\tilde{\pi}^{(k)}_*] = 0$. So ideally we would want our biasing $\tilde{\pi}^{(k)}$ to be as close to $\tilde{\pi}^{(k)}_*$ as possible. Similarly to MF, once the biasing density $\tilde{\pi}^{(k)}$ is chosen, the amount of variance reduction $c^{(k)}_{\mathrm{IS}}$ is fixed for any computational budget p.

2.4. Information reuse estimators

Information reuse is an uncertainty quantification concept develop to take advantage of the outer-loop context [33] by leveraging previous iteration HFMs as a control variate to the instance of the HFM at the current iteration. The information reuse (IR) estimator introduced in [33] uses the instance $F^{(k-1)}$ of the HFM corresponding to outer-loop variable $\boldsymbol{\lambda}^{(k-1)}$ at the previous outer-loop iteration $k-1$ as the control variate at iteration k: with a budget of p HFM evaluations and samples $\xi_1, \ldots, \xi_p$ drawn i.i.d. from $\pi^{(k)}$, the IR estimator is

$$\widehat{F}^{(k)}_{\mathrm{IR},p} = \left(\frac{1}{p/2} \sum_{i=1}^{p/2} F^{(k)}(\xi_i) \right) + \gamma \left(\widehat{F}^{(k-1)}_{\mathrm{IR},p} - \frac{1}{p/2} \sum_{i=1}^{p/2} F^{(k-1)}(\xi_i) \right), \qquad \widehat{F}^{(0)}_{\mathrm{IR},p} = \widehat{F}^{(0)}_{\mathrm{MC},p},$$

where γ is a constant which is chosen to minimize the MSE of the estimator. The kth IR estimator reuses both the IR estimator at iteration $k-1$ as well as the $(k-1)$st instance of the HFM.

Information reuse can also be combined with the multifidelity approach [33, 54] by replacing the MC estimators present in the IR estimator with MF estimators instead. The multifidelity information reuse (MFIR) estimator, using surrogates $G^{(k)}$ and $G^{(k-1)}$ is then

$$\widehat{F}^{(k)}_{\mathrm{MFIR},p} = \widehat{F}_{\mathrm{MF},F^{(k)},G^{(k)},\pi^{(k)},p/2} + \gamma\left(\widehat{F}^{(k-1)}_{\mathrm{MFIR},p} - \widehat{F}_{\mathrm{MF},F^{(k-1)},G^{(k-1)},\pi^{(k)},p/2}\right)$$

where the MFIR estimator is initialized with an MF estimator, $\widehat{F}^{(0)}_{\mathrm{MFIR},p} = \widehat{F}^{(0)}_{\mathrm{MF},p}$ and samples drawn from $\pi^{(k)}$ are the same across both MF estimators. The kth MFIR estimator depends on the $(k-1)$st MFIR estimator, the $(k-1)$st HFM, and the surrogates at the kth and $(k-1)$st iterations. Similarly to IR, the constant γ can be chosen to minimize the MSE of the estimator and depends on the variance of the $(k-1)$st MFIR estimator.

3. Adaptive information reuse estimators

In this section we discuss our adaptation of the original information reuse and multifidelity information reuse estimators [33, 54] to the case in which the input distribution depends on the outer-loop variables and thus evolves between iterations. In the case of a varying input distribution, the IR and MFIR estimators are biased with respect to $\mathbb{E}_{\pi^{(k)}}[F^{(k)}]$. This can be seen from the very first IR iteration:

$$\mathbb{E}_{\pi^{(1)}}[\widehat{F}^{(1)}_{\mathrm{IR},p}] = \mathbb{E}_{\pi^{(1)}}[F^{(1)}] + \gamma(\mathbb{E}_{\pi^{(0)}}[F^{(0)}] - \mathbb{E}_{\pi^{(1)}}[F^{(0)}])\,,$$

and likewise for the MFIR estimator. Key of the proposed adaptive MFIR (AMFIR) estimator is that it corrects the bias incurred by the evolution of the input distribution.

3.1. Adaptive information reuse

3.1.1. Adaptive information reuse (AIR) estimator

We introduce the adaptive information reuse (AIR) estimator with budget p as

$$\widehat{F}^{(k)}_{\mathrm{AIR},p} = \left(\frac{1}{p/2}\sum_{i=1}^{p/2} F^{(k)}(\xi_i)\right) + \gamma\left(\widehat{F}^{(k-1)}_{\mathrm{AIR},p} - \frac{1}{p/2}\sum_{i=1}^{p/2} F^{(k-1)}(\xi_i)W_k(\xi_i)\right)\,, \tag{6}$$

where $k = 1, 2, \ldots$ is the current outer-loop iteration and $\widehat{F}^{(0)}_{\mathrm{AIR},p} = \widehat{F}^{(0)}_{\mathrm{MC},p}$. The samples $\xi_1, \ldots, \xi_p$ are drawn i.i.d. from $\pi^{(k)}$. The weight $W_k(\xi)$ is

$$W_k(\xi) = \frac{\pi^{(k-1)}(\xi)}{\pi^{(k)}(\xi)}\,. \tag{7}$$

We now prove that this estimator is indeed unbiased for all k so long as $\widehat{F}^{(0)}_{\mathrm{AIR},p}$ is unbiased in the sense that $\mathbb{E}_{\pi^{(0)}}[\widehat{F}^{(0)}_{\mathrm{AIR},p}] = \mathbb{E}_{\pi^{(0)}}[F^{(0)}]$.

Proposition 3.1. *Suppose we initialize the AIR estimator with a MC estimator using a budget of p HFM evaluations, $\widehat{F}^{(0)}_{AIR,p} = \widehat{F}^{(0)}_{MC,p}$. Then, for the iteratively defined estimator $\widehat{F}^{(k)}_{AIR,p}$, it holds that $\mathbb{E}_{\pi^{(k)}}[\widehat{F}^{(k)}_{AIR,p}] = \mathbb{E}_{\pi^{(k)}}[F^{(k)}]$, which means $\widehat{F}^{(k)}_{AIR,p}$ is an unbiased estimator of $\mathbb{E}_{\pi^{(k)}}[F^{(k)}]$.*

Proof. The proof is by induction. Since the MC estimator is unbiased and the AIR estimator is initialized as an MC estimator, we have that $\mathbb{E}_{\pi^{(0)}}[\widehat{F}^{(0)}_{\mathrm{AIR},p}] = \mathbb{E}_{\pi^{(0)}}[F^{(0)}]$. For $k > 0$, suppose that $\mathbb{E}_{\pi^{(k-1)}}[\widehat{F}^{(k-1)}_{\mathrm{AIR},p}] =$

$\mathbb{E}_{\pi^{(k-1)}}[F^{(k-1)}]$. By the law of conditional expectation:

$$\begin{aligned}
\mathbb{E}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\mathrm{AIR},p}] &= \mathbb{E}_{\pi^{(k-1)}}[\mathbb{E}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\mathrm{AIR},p}|\widehat{F}^{(k-1)}_{\mathrm{AIR},p}]] = \mathbb{E}_{\pi^{(k-1)}}\left[\mathbb{E}_{\pi^{(k)}}[F^{(k)}] + \gamma(\widehat{F}^{(k-1)}_{\mathrm{AIR},p} - \mathbb{E}_{\pi^{(k)}}[F^{(k-1)}W_k])\right] \\
&= \mathbb{E}_{\pi^{(k-1)}}\left[\mathbb{E}_{\pi^{(k)}}[F^{(k)}] + \gamma\left(\widehat{F}^{(k-1)}_{\mathrm{AIR},p} - \mathbb{E}_{\pi^{(k)}}\left[F^{(k-1)}\frac{\pi^{(k-1)}}{\pi^{(k)}}\right]\right)\right] \\
&= \mathbb{E}_{\pi^{(k-1)}}\left[\mathbb{E}_{\pi^{(k)}}[F^{(k)}] - \gamma(\widehat{F}^{(k-1)}_{\mathrm{AIR},p} - \mathbb{E}_{\pi^{(k-1)}}[F^{(k-1)}])\right] \\
&= \mathbb{E}_{\pi^{(k)}}[F^{(k)}] - \gamma(\mathbb{E}_{\pi^{(k-1)}}[\widehat{F}^{(k-1)}_{\mathrm{AIR},p}] - \mathbb{E}_{\pi^{(k-1)}}[F^{(k-1)}]) = \mathbb{E}_{\pi^{(k)}}[F^{(k)}]
\end{aligned}$$

where the last step uses the inductive hypothesis that $\mathbb{E}_{\pi^{(k-1)}}[\widehat{F}^{(k-1)}_{\mathrm{AIR},p}] = \mathbb{E}_{\pi^{(k-1)}}[F^{(k-1)}]$. Thus we deduce that $\mathbb{E}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\mathrm{AIR},p}] = \mathbb{E}_{\pi^{(k)}}[F^{(k)}]$ holds for all k, and the AIR estimator is unbiased. □

3.1.2. Selecting balancing parameter γ in the AIR estimator

The γ that minimizes the variance of the AIR estimator can be derived analogously to the case of the IR estimator with an outer-loop iteration-independent input distribution. However, the optimal γ depends on the iteration k, which means it would need to be re-estimated in each outer-loop iteration in practice. Instead, we derive a γ^*_{AIR} that minimizes the asymptotic variance of the AIR estimator for $k \to \infty$. The following proposition derives the variance of the AIR estimator for $k \to \infty$.

3.1.3. Variance of AIR estimator

The γ that minimizes the variance of the AIR estimator can be derived analogously to the case of the IR estimator with a fixed input distribution [33]. This optimal γ^*_{AIR} is

$$\gamma^*_{\mathrm{AIR}} = \left(\frac{\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k)}{1+\eta_k}\right)\sqrt{\frac{\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]}{\mathrm{Var}_{\pi^{(k)}}[F^{(k-1)}W_k]}}, \qquad \eta_k = \frac{\mathrm{Var}_{\pi^{(k-1)}}[\widehat{F}^{(k-1)}_{\mathrm{AIR},p}]}{\mathrm{Var}_{\pi^{(k)}}[F^{(k-1)}W_k]/(p/2)}$$

and the variance of the AIR estimator with γ^*_{AIR} is

$$\mathrm{Var}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\mathrm{AIR},p}] = c^{(k)}_{\mathrm{AIR}}\frac{\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]}{p}, \qquad c^{(k)}_{\mathrm{AIR}} = 2\left(1 - \frac{\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k)^2}{1+\eta_k}\right) \tag{8}$$

where $c^{(k)}_{\mathrm{AIR}}$ is the variance reduction using the AIR estimator compared to MC. However, the optimal γ^* depends via η_k on the

3.1.4. Asymptotic variance of AIR

We now derive an asymptotic variance of the AIR estimator, which is independent of the We see that to compute the optimal choice of γ^*_{AIR}, we need to have an estimate of the variance of the $(k-1)$ AIR estimator. Since we initialize with an MC estimator, we can use samples from initially estimating $\mathbb{E}_{\pi^{(0)}}[F^{(0)}]$ to approximate $\mathrm{Var}_{\pi^{(0)}}[\widehat{F}^{(0)}_{\mathrm{AIR},p}] = \mathrm{Var}_{\pi^{(0)}}[\widehat{F}^{(0)}_{\mathrm{MC},p}]$. Then, as the outer-loop runs, we can use equation (8) to iteratively estimate $\mathrm{Var}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\mathrm{MC},p}]$.

Under some mild assumptions on the variances of $F^{(k)}$, $F^{(k-1)}W_k$, and their correlation under $\pi^{(k)}$ we can derive an asymptotic formula for the variance of the AIR estimator.

Proposition 3.2. *Suppose that variance reduction $c^{(k)}_{AIR}$ using the AIR estimator converges to $c_{AIR} \geq 0$ as $k \to \infty$. Moreover, suppose that $\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k)^2 \to \rho^2$ as $k \to \infty$ so that the correlation between current high-fidelity and reweighted previous high-fidelity converges. Finally, suppose that the ratio of variances $\mathrm{Var}_{\pi^{(k-1)}}[F^{(k-1)}]/\mathrm{Var}_{\pi^{(k)}}[F^{(k-1)}W_k] \to 1$ as $k \to \infty$. Then the asymptotic variance reduction is given by $c_{AIR} = 2\sqrt{1-\rho^2}$.*

Proof. Under the assumption that $\mathrm{Var}_{\pi^{(k-1)}}[F^{(k-1)}]/\mathrm{Var}_{\pi^{(k)}}[F^{(k-1)}W_k] \to 1$ as $k \to \infty$, we see that

$$\eta_k = \frac{\mathrm{Var}_{\pi^{(k-1)}}[\widehat{F}^{(k-1)}_{\mathrm{AIR},p}]}{\mathrm{Var}_{\pi^{(k)}}[F^{(k-1)}W_k]/(p/2)} = \frac{c^{(k-1)}_{\mathrm{AIR}}\,\mathrm{Var}_{\pi^{(k-1)}}[F^{(k-1)}]/p}{\mathrm{Var}_{\pi^{(k)}}[F^{(k-1)}W_k]/(p/2)} \to \frac{c_{\mathrm{AIR}}}{2} \quad \text{as } k \to \infty$$

Dividing the equation for $c_{\text{AIR}}^{(k)}$ (8) by 2, and passing to the limit we get

$$\frac{c_{\text{AIR}}^{(k)}}{2} = 1 - \frac{\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k)^2}{1+\eta_k} \quad \Longrightarrow \quad \frac{c_{\text{AIR}}}{2} = 1 - \frac{\rho^2}{1+(c_{\text{AIR}}/2)}$$

This system can then be solved,

$$\left(\frac{c_{\text{AIR}}}{2} - 1\right)\left(\frac{c_{\text{AIR}}}{2} + 1\right) = -\rho^2, \quad \Longrightarrow \quad \frac{c_{\text{AIR}}^2}{4} = 1 - \rho^2$$
$$\Longrightarrow \quad c_{\text{AIR}} = 2\sqrt{1-\rho^2}$$

□

The results of Proposition 3.2 motivates approximating the variance reduction of the AIR estimator at an iteration k as

$$\hat{c}_{\text{AIR}}^{(k)} = 2\sqrt{1 - \rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k)^2} \tag{9}$$

where $|\hat{c}_{\text{AIR}}^{(k)} - c_{\text{AIR}}^{(k)}| \to 0$ as $k \to \infty$. A major advantage of $\hat{c}_{\text{AIR}}^{(k)}$ is that unlike equation (8), it does not directly depend on the estimation done at the previous step. That is, while the AIR estimator at the kth iteration requires the AIR estimator at the $(k-1)$st iteration as well as an estimate of its variance, $\hat{c}_{\text{AIR}}^{(k)}$ depends only on the correlation $\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k)$ which can be estimated independently of any estimation done at the $(k-1)$st iteration.

Remark 3.3. *The assumptions required in deriving $\hat{c}_{AIR}^{(k)}$ are somewhat natural, as we assume the outer-loop variables $\boldsymbol{\lambda}^{(k)}$ to converge to $\boldsymbol{\lambda}$ as $k \to \infty$. Assuming $\pi_{\boldsymbol{\lambda}^{(k)}}^{(k)}$ and $F_{\boldsymbol{\lambda}^{(k)}}^{(k)}$ depend continuously on $\boldsymbol{\lambda}^{(k)}$, then $\pi_{\boldsymbol{\lambda}^{(k)}}^{(k)}$ will converge to a steady-state distribution and $F_{\boldsymbol{\lambda}^{(k)}}^{(k)}$ will converge to a steady-state HFM as $k \to \infty$, where the modes of convergence will depend on properties of $\pi^{(k)}$ and $F^{(k)}$. Subsequently, this will yield convergence of the necessary correlations and variance ratios.*

Remark 3.4. *At each iteration $k = 1, 2, \ldots$, $\hat{c}_{AIR}^{(k)}$ only depends on $F^{(k)}$, $F^{(k-1)}$, $\pi^{(k)}$, and $\pi^{(k-1)}$, and thus can be readily estimated using the same samples which built the AIR estimator itself. This allows us to monitor the efficacy of the AIR estimator compared to other estimators during the outer-loop. The variance reduction using the AIR estimator will be large when the correlation between the HFMs $F^{(k)}$ and $F^{(k-1)}W_k$ under $\pi^{(k)}$ is close to 1.*

3.2. Multifidelity adaptive information reuse

Correcting the bias in the MFIR estimator for dynamically changing $\pi^{(k)}$ is analogous to the correction of the AIR estimator. Namely, we use $F^{(k-1)}W_k$ as the HFM control variate instead of $F^{(k-1)}$.

3.2.1. Multifidelity adaptive information reuse estimator (MFAIR)

We define the multifidelity adaptive information reuse estimator (MFAIR) estimator with a computational budget of p equivalent HFM evaluations is then

$$\widehat{F}_{\text{MFAIR},p}^{(k)} = \widehat{F}_{\text{MF},F^{(k)},G^{(k)},\pi^{(k)},p/2} + \gamma\left(\widehat{F}_{\text{MFAIR},p}^{(k-1)} - \widehat{F}_{\text{MF},F^{(k-1)}W_k,G^{(k-1)},\pi^{(k)},p/2}\right) \tag{10}$$

where we initialize the MFAIR estimator with an MF estimator, $\widehat{F}_{\text{MFAIR},p}^{(0)} = \widehat{F}_{\text{MF},p}^{(0)}$. Similarly to the AIR estimator, it is necessary that the samples ξ drawn i.i.d. from $\pi^{(k)}$ in $\widehat{F}_{\text{MF},F^{(k)},G^{(k)},\pi^{(k)},p/2}$ are the same as those drawn in $\widehat{F}_{\text{MF},F^{(k-1)}W_k,G^{(k-1)},\pi^{(k)},p/2}$. This is needed to ensure that the samples $F^{(k)}(\xi_i)$, $G^{(k)}(\xi_i)$, $F^{(k-1)}(\xi_i)W_k(\xi_i)$, $G^{(k-1)}(\xi_i)$ will all be correlated.

3.2.2. Unbiasedness and variance of MFAIR

The proof that the MFAIR estimator is unbiased for all k follows immediately from that of Proposition 3.1, as the MF estimator is unbiased and the same induction argument holds. Since the MFAIR estimator is unbiased, the MSE is equivalent to the variance, and the optimal choice of γ can again be found by minimizing this variance with respect to γ using first-order information. The derivation of the optimal γ^*_{MFAIR} and the variance of the subsequent MFAIR estimator are analogous to the case of MFIR with fixed input distribution [33]. The optimal γ^*_{MFAIR} is given by

$$\gamma^*_{\text{MFAIR}} = \frac{\text{COV}_k}{\text{Var}_{\pi^{(k-1)}}[\widehat{F}^{(k-1)}_{\text{MFAIR},p}] + \text{Var}_{\pi^{(k)}}[\widehat{F}_{\text{MF},F^{(k-1)}W_k,G^{(k-1)},\pi^{(k)},p/2}]}$$

where the term COV_k is the covariance of $\widehat{F}_{\text{MF},F^{(k)},G^{(k)},\pi^{(k)},p/2}$ and $\widehat{F}_{\text{MF},F^{(k-1)}W_k,G^{(k-1)},\pi^{(k)},p/2}$ under the distribution $\pi^{(k)}$, given by

$$\begin{aligned}\text{COV}_k =& \Bigg[\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k) + \left(1 - \frac{1}{r^*_k}\right)\rho_{\pi^{(k)}}(F^{(k)}, G^{(k)})\rho_{\pi^{(k)}}(F^{(k-1)}W_k, G^{(k-1)})\rho_{\pi^{(k)}}(G^{(k)}, G^{(k-1)}) \\ &- \rho_{\pi^{(k)}}(F^{(k)}, G^{(k)})\rho_{\pi^{(k)}}(F^{(k-1)}W_k, G^{(k)}) - \rho_{\pi^{(k)}}(F^{(k-1)}W_k, G^{(k-1)})\rho_{\pi^{(k)}}(F^{(k)}, G^{(k-1)})\Bigg] \\ &\times \frac{\sqrt{\text{Var}_{\pi^{(k)}}[F^{(k)}]\,\text{Var}_{\pi^{(k)}}[F^{(k-1)}W_k]}}{n}\end{aligned} \tag{11}$$

where

$$r^*_k = \frac{m}{n} = \min\left(\sqrt{\frac{w\rho_{\pi^{(k)}}(F^{(k)}, G)^2}{1 - \rho_{\pi^{(k)}}(F^{(k)}, G)^2}}, \sqrt{\frac{w\rho_{\pi^{(k)}}(F^{(k-1)}W_k, G)^2}{1 - \rho_{\pi^{(k)}}(F^{(k-1)}W_k, G)^2}}\right)$$

and where n is the number of HFM evaluations m is the number of low-fidelity model evaluations, satisfying $p/2 = n + m/w$.

The variance of the MFAIR estimator using optimal γ^*_{MFAIR} can be given by:

$$\text{Var}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\text{MFAIR},p}] = \text{Var}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\text{MF},p/2}] - \frac{\text{COV}_k^2}{\text{Var}_{\pi^{(k-1)}}[\widehat{F}^{(k-1)}_{\text{MFAIR},p}] + \text{Var}_{\pi^{(k)}}[\widehat{F}_{\text{MF},F^{(k-1)}W_k,G^{(k-1)},\pi^{(k)},p/2}]} \tag{12}$$

3.2.3. Asymptotic variance of MFAIR

We now derive a new asymptotic expression for the variance reduction of the MFAIR estimator compared to the MC estimator, for the case where $G^{(k)} = G$ for all k. That is, in the case where the surrogate model is the same for the entire outer-loop process.

Let $c^{(k)}_{\text{MFAIR}}$ be the variance reduction using the MFAIR estimator compared to the MC estimator, so that $\text{Var}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\text{MFAIR},p}] = c^{(k)}_{\text{MFAIR}}\,\text{Var}_{\pi^{(k)}}[F^{(k)}]/p$.

Proposition 3.5. *Consider the MFAIR estimator with $G^{(k)} = G$ for all k, and suppose that $c^{(k)}_{MFAIR} \to c_{MFAIR} \geq 0$ as $k \to \infty$. Additionally suppose that the correlations between HFMs and surrogate models converge to the same value, so that $\rho_{\pi^{(k)}}(F^{(k)}, G) \to \rho$ as $k \to \infty$ and $\rho_{\pi^{(k)}}(F^{(k-1)}W_k, G) \to \rho$ as $k \to \infty$ for $\rho < 1$. Likewise, suppose that the correlation between the current and previous HFMs converge so that $\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k) \to \hat{\rho}$ as $k \to \infty$. Lastly assume that $\text{Var}_{\pi^{(k-1)}}[F^{(k-1)}]/\,\text{Var}_{\pi^{(k)}}[F^{(k-1)}W_k] \to 1$ as $k \to \infty$. Then*

$$c_{MFAIR} = 2\left(1 + \frac{r^*}{w}\right)\sqrt{(1 - \varphi^*\rho^2)^2 - (\hat{\rho} - \varphi^*\rho^2)^2}$$

where $r^* = \lim_{k\to\infty} r^*_k$, and $\varphi^* = 1 - (1/r^*)$.

Proof. Let $V_k = c^{(k)}_{\text{MFAIR}}/[2(1+(r^*_k/w))]$, so that $\text{Var}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\text{MFAIR},p}] = V_k \,\text{Var}_{\pi^{(k)}}[F^{(k)}]/n$. We can express the variance of the two MF estimators [33] within the MFAIR estimator as

$$\text{Var}_{\pi^{(k)}}[\widehat{F}^{(k)}_{\text{MF},p/2}] = \left[1 - \left(1 - \frac{1}{r^*_k}\right)\rho_{\pi^{(k)}}(F^{(k)}, G)^2\right] \frac{\text{Var}_{\pi^{(k)}}[F^{(k)}]}{n}$$

$$\text{Var}_{\pi^{(k)}}[\widehat{F}_{\text{MF},F^{(k-1)}W_k,G^{(k-1)},\pi^{(k)},p/2}] = \left[1 - \left(1 - \frac{1}{r^*_k}\right)\rho_{\pi^{(k)}}(F^{(k-1)}W_k, G)^2\right] \frac{\text{Var}_{\pi^{(k)}}[F^{(k-1)}W_k]}{n}.$$

Define $\varphi^*_k = 1 - (1/r^*_k)$. Then based on our assumptions of convergence, $\varphi^*_k \to \varphi^*$ for some $\varphi^* > 0$. To see that φ^* is small, we note that r^*_k is the ratio of surrogate model evaluations to HFM evaluations. Assuming the correlation remains large between the HFMs and surrogate, it is expected that r^*_k is large, and therefore φ^*_k is small.

Since we are using $G^k = G$ for all k, the covariance term can be written as:

$$\text{COV}_k = \left(\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k) - \varphi^*_k \rho_{\pi^{(k)}}(F^{(k)}, G)\rho_{\pi^{(k)}}(F^{(k-1)}W_k, G)\right) \frac{\sqrt{\text{Var}_{\pi^{(k)}}[F^{(k)}]\,\text{Var}_{\pi^{(k)}}[F^{(k-1)}W_k]}}{n}$$

Plugging the covariance term and MF variance equations into equation (12) and dividing both sides by $\text{Var}_{\pi^{(k)}}[F^{(k)}]/n$, we get

$$V_k = \left(1 - \varphi^*_k \rho_{\pi^{(k)}}(F^{(k)}, G)^2\right) - \frac{\left(\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k) - \varphi^*_k \rho_{\pi^{(k)}}(F^{(k)}, G)\rho_{\pi^{(k)}}(F^{(k-1)}W_k, G)\right)^2}{V_{k-1}\left(\text{Var}_{\pi^{(k-1)}}[F^{(k-1)}]/\,\text{Var}_{\pi^{(k)}}[F^{(k-1)}W_k]\right) + (1 - \varphi^*_k \rho_{\pi^{(k)}}(F^{(k-1)}W_k, G)^2)}$$

Passing to the limit we get

$$V = (1 - \varphi^*\rho^2) - \frac{(\hat{\rho} - \varphi^*\rho^2)^2}{V + (1 - \varphi^*\rho^2)}$$

Simplifying and solving for V we get

$$(V - (1 - \varphi^*\rho^2))(V + (1 - \varphi^*\rho^2)) = -(\hat{\rho} - \varphi^*\rho^2) \quad \Longrightarrow \quad V = \sqrt{(1 - \varphi^*\rho^2)^2 - (\hat{\rho} - \varphi^*\rho^2)^2}$$

Under our assumptions, we will also have $r^*_k \to r^*$ and thus

$$c_{\text{MFAIR}} = 2\left(1 + \frac{r^*}{w}\right)\sqrt{(1 - \varphi^*\rho^2)^2 - (\hat{\rho} - \varphi^*\rho^2)^2}$$

□

Analogous to the case for the AIR estimator, Proposition 3.5 motivates approximating the variance reduction using the MFAIR estimator at an iteration k by

$$\hat{c}^{(k)}_{\text{MFAIR}} = 2\left(1 + \frac{r^*_k}{w}\right)\sqrt{\left(1 - \varphi^*_k \rho_{\pi^{(k)}}(F^{(k)}, G)^2\right)^2 - \left(\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k) - \varphi^*_k \rho_{\pi^{(k)}}(F^{(k)}, G)^2\right)^2} \tag{13}$$

where $|\hat{c}^{(k)}_{\text{MFAIR}} - c^{(k)}_{\text{MFAIR}}| \to 0$ as $k \to \infty$. As for $\hat{c}^{(k)}_{\text{AIR}}$ for the AIR estimator, $\hat{c}^{(k)}_{\text{MFAIR}}$ can be estimated independently of any estimation done at the previous iteration.

We can see that the variance reduction using the MFAIR estimator is large when the correlation between HFMs $\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)W_k})$ is large. Moreover, we see that this variance reduction will be even larger if the correlation between the high-fidelity and surrogate models $\rho_{\pi^{(k)}}(F^{(k)}, G)$ is also large. Since $\hat{c}^{(k)}_{\text{MFAIR}}$ can be measured during the outer-loop, this allows us to monitor the efficacy of the MFAIR estimator compared to other estimators.

4. Boosting variance reduction with meta estimators

The MF, IS, and AIR methods all provide variance reduction by taking advantage of different features of the problem at hand. Therefore, various combinations of these methods have the benefit of simultaneously taking advantage of some or all these features. In this section we detail three new combinations of variance reduction techniques, namely combining IS estimators with MF estimators, IS estimators with AIR estimators, and a triple combined estimator which combines the MF, IS, and AIR estimators. For each method we derive the variance of these estimators as well as asymptotic approximations for the two methods which contain information reuse.

4.1. Importance sampled multifidelity method

To combine the IS estimator with the MF estimator, we simply replace the MC estimators present in the MF estimator (2) with IS estimators instead. Similarly to the reweighted HFM $\widetilde{F}^{(k)}$, let $\widetilde{G}^{(k)} = G^{(k)}\pi^{(k)}/\tilde{\pi}^{(k)}$ be the reweighted surrogate model, where we are using importance weights $\pi^{(k)}/\tilde{\pi}^{(k)}$. The importance sampled multifidelity (ISMF) estimator with computational budget equivalent to p HFM evaluations is

$$\widehat{F}^{(k)}_{\mathrm{ISMF},p} = \widehat{F}_{\mathrm{MF},\widetilde{F}^{(k)},\widetilde{G}^{(k)},\tilde{\pi}^{(k)},p} \tag{14}$$

where now the samples ξ_i are drawn from $\tilde{\pi}^{(k)}$ instead of π. The variance of the ISMF estimator is given by

$$\mathrm{Var}_{\tilde{\pi}^{(k)}}\left[\widehat{F}^{(k)}_{\mathrm{ISMF},p}\right] = \left(\sqrt{1-\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)},\widetilde{G}^{(k)})^2} + \sqrt{\frac{\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)},\widetilde{G^{(k)}})^2}{\tilde{w}}}\right)^2 c^{(k)}_{\mathrm{IS}}\frac{\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]}{p} \tag{15}$$

using $\mathrm{Var}_{\tilde{\pi}^{(k)}}[\widetilde{F}^{(k)}] = c^{(k)}_{\mathrm{IS}}\,\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]$.

Note that the performance of this estimator now depends on the correlation between $\widetilde{F}^{(k)}$ and $\widetilde{G}^{(k)}$ under $\tilde{\pi}^{(k)}$ and depends on the cost ratio $\tilde{w}$ when samples are drawn from $\tilde{\pi}^{(k)}$. The latter difference can be substantial, depending on how the surrogate G is constructed. As a result, there may be a significant difference in the multifidelity cost ratio when sampling from $\tilde{\pi}^{(k)}$ instead of $\pi^{(k)}$. The difference in correlation $\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)},\widetilde{G}^{(k)})$ compared to $\rho_{\pi^{(k)}}(F^{(k)},G^{(k)})$ as in MF may also be noticeable, depending on the quality of the surrogate G and the biasing $\tilde{\pi}^{(k)}$.

Let $c_{\mathrm{MF}}(F,G,\pi)$ denote the variance reduction using the estimator $\widehat{F}_{\mathrm{MF},F,G,\pi,p}$ compared to the MC estimator of F with p samples. Then the variance reduction of the MF estimator is

$$c^{(k)}_{\mathrm{MF}} = c_{\mathrm{MF}}(F^{(k)},G^{(k)},\pi^{(k)})$$

In comparison, the variance reduction of the ISMF estimator is

$$c^{(k)}_{\mathrm{ISMF}} = c_{\mathrm{MF}}(\widetilde{F}^{(k)},\widetilde{G}^{(k)},\tilde{\pi}^{(k)}) \times c^{(k)}_{\mathrm{IS}}$$

Thus the ISMF estimator provides a quasi-multiplicative variance reduction, in that $c_{\mathrm{ISMF}^{(k)}}$ is the product of variance reduction from an MF estimator using $\widetilde{F}^{(k)}$, $\widetilde{G}^{(k)}$, $\tilde{\pi}^{(k)}$ instead of $F^{(k)}$, $G^{(k)}$, $\pi^{(k)}$, with the variance reduction of the IS estimator. So long as $c_{\mathrm{MF}}(\widetilde{F}^{(k)},\widetilde{G}^{(k)},\tilde{\pi}^{(k)})$ is not too large compared to $c_{\mathrm{MF}}(F^{(k)},G^{(k)},\pi^{(k)})$, we can possibly expect ISMF to outperform both the MF and IS estimators on their own.

4.2. Importance sampled adaptive information reuse

To combine IS with AIR we just replace the MC estimators present in the AIR estimator (6) with IS estimators instead. For a budget of p HFM evaluations, we define the importance sampled adaptive information reuse (ISAIR) estimator as

$$\widehat{F}^{(k)}_{\mathrm{ISAIR},p} = \left(\frac{1}{p/2}\sum_{i=1}^{p/2}\widetilde{F}^{(k)}(\xi_i)\right) + \gamma\left(\widehat{F}^{(k-1)}_{\mathrm{ISAIR},p} - \frac{1}{p/2}\sum_{i=1}^{p/2}\widetilde{F}^{(k-1)}(\xi_i)W_k(\xi_i)\right) \tag{16}$$

where now ξ_i are drawn i.i.d. from $\tilde{\pi}^{(k)}$. The optimal $\gamma^*_{\mathrm{ISAIR}}$ and the variance of the ISAIR estimator are analogous to that of the AIR estimator, but with $\tilde{\pi}^{(k)}$ replacing $\pi^{(k)}$ and with all HFMs replaced with their importance weighted counterparts. The optimal $\gamma^*_{\mathrm{ISAIR}}$ is then

$$\gamma^*_{\mathrm{ISAIR}} = \left(\frac{\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)},\widetilde{F}^{(k-1)}W_k)}{1+\tilde{\eta}_k}\right)\sqrt{\frac{\mathrm{Var}_{\tilde{\pi}^{(k)}}[\widetilde{F}^{(k)}]}{\mathrm{Var}_{\tilde{\pi}^{(k)}}[\widetilde{F}^{(k-1)}W_k]}}, \qquad \tilde{\eta}_k = \frac{\mathrm{Var}_{\tilde{\pi}^{(k-1)}}[\widehat{F}^{(k-1)}_{\mathrm{ISIR},p}]}{\mathrm{Var}_{\tilde{\pi}^{(k)}}[\widetilde{F}^{(k-1)}W_k]/(p/2)}$$

and the variance of the ISAIR estimator using $\gamma^*_{\mathrm{ISAIR}}$ is

$$\mathrm{Var}_{\tilde{\pi}^{(k)}}[\widehat{F}^{(k)}_{\mathrm{ISIR},p}] = c^{(k)}_{\mathrm{ISAIR}}\frac{\mathrm{Var}_{\pi^{(k)}}[F^{(k)}]}{p}, \qquad c^{(k)}_{\mathrm{ISAIR}} = 2\left(1-\frac{\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)},\widetilde{F}^{(k-1)}W_k)^2}{1+\tilde{\eta}_k}\right)c^{(k)}_{\mathrm{IS}} \tag{17}$$

Similar to the AIR estimator, we can derive an approximation for the asymptotic ISAIR variance reduction analogous to (9). The derivation of this approximation follows from the proof of Proposition 3.2 but with the necessary models and distributions substituted, assuming that the variance reduction from IS alone $c_{\text{IS}}^{(k)}$ also converges as $k \to \infty$.

$$\hat{c}_{\text{ISAIR}}^{(k)} = \left(2\sqrt{1 - \rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)}, \widetilde{F}^{(k-1)}W_k)^2}\right) c_{\text{IS}}^{(k)} \tag{18}$$

where $|\hat{c}_{\text{ISAIR}}^{(k)} - c_{\text{ISAIR}}^{(k)}| \to 0$ as $k \to \infty$.

Let $c_{\text{AIR}}(\{F^{(l)}, \pi^{(l)}\}_{l=0}^k)$ denote the variance reduction using the AIR estimator with a history of HFMs and input distributions $F^{(l)}, \pi^{(l)}$ for $l = 0, \ldots, k$, so that

$$c_{\text{AIR}}^{(k)} = c_{\text{AIR}}(\{F^{(l)}, \pi^{(l)}\}_{l=0}^k)$$

Then the variance reduction of the ISAIR estimator is

$$c_{\text{ISAIR}}^{(k)} = c_{\text{AIR}}(\{\widetilde{F}^{(l)}, \tilde{\pi}^{(l)}\}_{l=0}^k) \times c_{\text{IS}}^{(k)}$$

So analogously to the ISMR estimator, the ISAIR estimator provides a pseuedo-multiplicative variance reduction, in that $c_{\text{ISAIR}}^{(k)}$ is the product of variance reduction from an AIR estimator using $\widetilde{F}^{(l)}, \tilde{\pi^{(l)}}$ instead of $F^{(l)}, \pi^{(l)}$ for $l = 0, \ldots, k$, with the variance reduction of the IS estimator. Assuming $c_{\text{AIR}}(\{\widetilde{F}^{(l)}, \tilde{\pi}^{(l)}\}_{l=0}^k)$ is not significantly larger than $c_{\text{AIR}}^{(k)}$, we can expect ISAIR to outperform both AIR and IS individually.

4.3. Importance sampled multifidelity adaptive information reuse estimator

Lastly, we detail an estimator which simultaneously combines the MF, IS, and AIR method. To do this, we replace the MC estimators present in the AIR estimator (6) with ISMF estimators. Equivalently, this can be viewed as introducing importance weights into the MFAIR scheme. For an equivalent budget of p HFM samples, we define the importance sampled multifidelity adaptive information reuse estimator (ISMFAIR) as

$$\widehat{F}_{\text{ISMFAIR},p}^{(k)} = \widehat{F}_{\text{MF},\widetilde{F}^{(k)},\widetilde{G}^{(k)},\tilde{\pi}^{(k)},p/2} + \gamma\left(\widehat{F}_{\text{ISMFAIR},p}^{(k-1)} - \widehat{F}_{\text{MF},\widetilde{F}^{(k-1)}W_k,\widetilde{G}^{(k-1)},\tilde{\pi}^{(k)},p/2}\right) \tag{19}$$

Just as for the ISAIR estimator, the optimal $\gamma_{\text{ISMFAIR}}^*$ and variance of the ISMFAIR estimator are identical to that of the MFAIR estimator but with the relevent substitutions made. The optimal $\gamma_{\text{ISMFAIR}}^*$ is given by

$$\gamma_{\text{ISMFAIR}}^* = \frac{\widetilde{\text{COV}}_k}{\text{Var}_{\tilde{\pi}^{(k-1)}}[\widehat{F}_{\text{ISMFAIR},p}^{(k-1)}] + \text{Var}_{\tilde{\pi}^{(k)}}[\widehat{F}_{\text{ISMFAIR},p}^{(k-1)} - \widehat{F}_{\text{MF},\widetilde{F}^{(k-1)}W_k,\widetilde{G}^{(k-1)},\tilde{\pi}^{(k)},p/2}]}$$

where the term $\widetilde{\text{COV}}_k$ is the covariance of $\widehat{F}_{\text{MF},\widetilde{F}^{(k)},\widetilde{G}^{(k)},\tilde{\pi}^{(k)},p/2}$ and $\widehat{F}_{\text{MF},\widetilde{F}^{(k-1)}W_k,\widetilde{G}^{(k-1)},\tilde{\pi}^{(k)},p/2}$ under the distribution $\tilde{\pi}^{(k)}$. The formula for $\widetilde{\text{COV}}_k$ is identical to the formula for COV_k, equation (11), but with all models replaced by their importance weighted counterparts and $\pi^{(k)}$ replace by $\tilde{\pi}^{(k)}$.

The variance of the optimal ISMFAIR estimator is

$$\text{Var}_{\tilde{\pi}^{(k)}}[\widehat{F}_{\text{ISMFAIR},p}^{(k)}] = \text{Var}_{\tilde{\pi}^{(k)}}[\widehat{F}_{\text{ISMFAIR},p/2}^{(k)}] - \frac{\widetilde{\text{COV}}_k^2}{\text{Var}_{\tilde{\pi}^{(k-1)}}[\widehat{F}_{\text{ISMFAIR},p}^{(k-1)}] + \text{Var}_{\tilde{\pi}^{(k)}}[\widehat{F}_{\text{ISMFAIR},p}^{(k-1)} - \widehat{F}_{\text{MF},\widetilde{F}^{(k-1)}W_k,\widetilde{G}^{(k-1)},\tilde{\pi}^{(k)},p/2}]}$$

Let $c_{\text{ISMFAIR}}^{(k)}$ be the variance reduction of the MFAIR estimator compared to the MC estimator so that $\text{Var}_{\tilde{\pi}^{(k)}}[\widehat{F}_{\text{ISMFAIR},p}^{(k)}] = c_{\text{ISMFAIR}}^{(k)} \text{Var}_{\pi^{(k)}}[F^{(k)}]/p$. Just as for the MFAIR estimator, if we consider $\widetilde{G}^{(k)} = \widetilde{G}$ for all k, we can derive the same approximation for the asymptotic variance reduction, where again the derivation is identical upon substituting the relevant functions

$$\hat{c}_{\text{ISMFAIR}}^{(k)} = \left(2\left(1 + \frac{\tilde{r}_k^*}{\tilde{w}}\right)\sqrt{\left(1 - \tilde{\varphi}_k^* \rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)}, \widetilde{G})^2\right)^2 - \left(\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)}, \widetilde{F}^{(k-1)}W_k) - \tilde{\varphi}_k^* \rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)}, \widetilde{G})^2\right)^2}\right) c_{\text{IS}}^{(k)} \tag{20}$$

where $\tilde{r}_k^* = m/n$ where m is the number of $\widetilde{G}$ evaluations and n is the number of $\widetilde{F}^{(k)}$ evaluations as dictated in the MF estimator. Subsequently, $\tilde{\varphi}_k^* = 1 - (1/\tilde{r}_k^*)$. As before, this approximation satisfies $|\hat{c}^{(k)}_{\text{ISMFAIR}} - c^{(k)}_{\text{ISMFAIR}}| \to 0$ as $k \to \infty$.

Let $c_{\text{MFAIR}}(\{F^{(l)}, G^{(l)}, \pi^{(l)}\}_{l=0}^k)$ denote the variance reduction using the MFAIR estimator with a history of HFMs, surrogates, and input distributions $F^{(l)}, G^{(l)}, \pi^{(l)}$ for $l = 0, \ldots, k$, so that

$$c^{(k)}_{\text{MFAIR}} = c_{\text{MFAIR}}(\{F^{(l)}, G^{(l)}, \pi^{(l)}\}_{l=0}^k)$$

Then the variance reduction of the ISMFAIR estimator is

$$c^{(k)}_{\text{ISMFAIR}} = c_{\text{MFAIR}}(\{\widetilde{F}^{(l)}, \widetilde{G}^{(l)}, \tilde{\pi}^{(l)}\}_{l=0}^k) \times c^{(k)}_{\text{IS}}$$

Thus, akin to our other meta estimators, the ISMFAIR estimator provides quasi-multiplicative variance reduction, in that $c^{(k)}_{\text{ISMFAIR}}$ is the product of variance reduction from a MFAIR estimator using $\widetilde{F}^{(l)}, \widetilde{G}^{(l)}, \tilde{\pi^{(l)}}$ instead of $F^{(l)}, G^{(l)}, \pi^{(l)}$ for $l = 0, \ldots, k$, with the variance reduction of the IS estimator. So long as $c_{\text{MFAIR}}(\{\widetilde{F}^{(l)}, \widetilde{G}^{(l)}, \tilde{\pi}^{(l)}\}_{l=0}^k)$ is not significantly larger than $c^{(k)}_{\text{MFAIR}}$, we may expect ISMFAIR to outperform the MFAIR and IS estimators on their own.

In the next section we detail our application of these methods for the specific problem of energetic particle confinement estimation during the stellarator optimization loop, as well as discuss our data-driven methodology for both the surrogate model G and biasing density $\pi^{(k)}$.

5. Energetic particles in stellarators

We now turn to the application of our framework to the construction of efficient estimators for energetic particle confinement in stellarators. We first introduce our physical model for the dynamics of energetic particles in fusion devices, and then present our construction of efficient estimators for the confining quality of the magnetic field during optimization.

5.1. Energetic particle dynamics and stellarator optimization

We consider the dynamics of 3.5 MeV alpha particles born as a result of deuterium-tritium fusion in a three-dimensional stellarator magnetic field. Fusion reactions may be approximated as a probabilistic process with the following properties: the direction of the velocity of alpha particles at birth follows a uniform distribution, and if the deuterium and tritium nuclei are at the same uniform temperature throughout the domain, then the location of birth of alpha particles in that domain also follows a uniform distribution. Studies of alpha particle confinement often rely on a Monte Carlo framework to properly capture the consequences of the probabilistic nature of the process [16, 18, 56, 19, 20, 29, 21]. Deterministic measures of energetic particle confinement have been proposed for the design of magnetic field with good confinement properties [22, 23, 24, 13], but their reliability and scope are limited [24, 13], due to the wide variety of particle orbits and of loss mechanisms [14, 57, 58, 13, 59, 11].

In the Monte Carlo approach, which is considered the most accurate, one randomly selects initial conditions for the alpha particles corresponding to the probabilistic birth process described above, and then numerically integrates their trajectories to determine the fraction of particles in the sample that is eventually lost, as well as the average confinement time for that sample. In principle, these trajectories should account for the effect of collisions with the thermalized electrons and deuterium and tritium ions [16, 58, 24, 60]. In practice however, ignoring the effect of collisions can still provide good accuracy for the energetic particle loss estimates, for times of flight up to a large fraction of the alpha particle slowing down time due to collisions [61, 24, 60]. Since energetic particle losses occurring before their characteristic slowing down time are the most detrimental for a fusion nuclear power plant [62, 58, 23, 24, 13], from both power balance and material damage perspectives, and since collisionless orbits are easier and less computationally expensive to integrate than orbits including collisions, many energetic particle confinement studies for stellarator optimization are done based on collisionless orbits [25, 13, 18, 19, 20, 29, 21]. This is also what we do in this article. We however stress that the methods we present here also apply to orbit integrators which are able to account for collisions. The analysis of the gains in efficiency provided by our methods when collisions are taken into account is left for future work.

The full collisionless dynamics of energetic particles born in stellarators is governed by the Lorentz force, according to Newton's second law of motion: $md\mathbf{v}/dt = q\mathbf{v} \times \mathbf{B}$, where m is the particle mass, q the particle charge, and $\mathbf{B}$ the magnetic field at the particle location. Computing the orbits given by these ordinary differential equations (ODEs) for the full extent of the slowing down time scale is computationally expensive, due to the multi-scale nature of the motion: to lowest order, the particles execute a fast quasi-helical motion centered on a magnetic field line, but the loss of confinement is due to the small departure of this motion from a perfect helix, which is called drift, and occurs on a much slower time scale [63, 25, 13]. To this day, in the absence of numerical ODE integrators capable of relying on the scale separation between the two types of motion to accelerate the computation of the particle orbits without a significant loss of accuracy, it remains unrealistic to include Monte Carlo simulations based on the full Newton's equations in stellarator optimization and design studies. To address this challenge, physicists have relied on a multiple time scale analysis to derive *guiding center equations* [25], which arise from averaging the equations of motion given by the Lorentz force over the fast helical motion [63, 64], and which describe the motion of the average particle location during its helical motion, called the guiding center. In the limit in which the radius of the particle helical motion is negligible compared to the typical length scale of variation of the magnetic field, the guiding center equations provide an excellent approximation of the exact particle motion [25, 24]. This regime, which is observed for strong magnetic fields, is the regime of interest for magnetic fusion reactors. The accuracy of the guiding-center orbits in that regime has been verified numerically [65]. Since the guiding center equations are much less computationally expensive to integrate than the full Newton's equations, they are most commonly used for stellarator optimization [66, 23, 24, 17, 18, 19, 20, 11, 21]. We therefore also apply our variance reduction framework to the guiding center equations in this work, and not to the full Newton's equations. It has been recognized that the guiding center equations may have limited accuracy in a few situations of interest for reactor design [24, 67]. We note that all the variance reduction methods discussed in this paper can also be applied to full orbit dynamics.

The guiding center equations can be represented as a four-dimensional system of ordinary differential equations (ODEs) corresponding to three spatial dimensions, and one dimension for the parallel velocity v. For the vacuum fields we will consider for our numerical tests in the next section, these dynamics are given by [68]

$$\dot{\mathbf{x}} = v\frac{\mathbf{B}^{(k)}}{B^{(k)}} + \frac{m}{q(B^{(k)})^3}\left(\frac{v_\perp^2}{2} + v^2\right)\mathbf{B}^{(k)} \times \nabla B^{(k)} \tag{21}$$

$$\dot{v} = -\frac{\mu}{B^{(k)}}\mathbf{B}^{(k)} \cdot \nabla B^{(k)} \tag{22}$$

where m is the particle mass, q is the particle charge, $\mathbf{x}$ is the position vector of the energetic particle, $\mathbf{B}^{(k)}$ is the magnetic field at the kth iteration determined by the outer-loop variables $\boldsymbol{\lambda}^{(k)}$, $B^{(k)} = |\mathbf{B}^{(k)}(x, y, z)|$ is the field strength, μ is the magnetic moment [63, 25], and $v_\perp^2 = 2\mu B^{(k)}$. Given an initial position (x_0, y_0, z_0) and initial parallel velocity v_0, energetic particles are traced by solving equations (21-22) until some final time $T_{\max}$, such as the characteristic alpha particle slowing down time, or until they reach a closed flux surface we label as the plasma edge, and are considered lost. We stress once more that our methods are not limited to dynamics given by equations (21-22). These equations are highlighted here because they correspond to the situations we considered for our numerical examples.

As is common in many confinement studies [16, 18, 19, 20, 21], we model particles as only being spawned on a single flux surface $S^{(k)}_{\text{spawn}}$ and consider particles lost when they reach a flux surface $S^{(k)}_{\text{exit}}$. The surfaces $S^{(k)}_{\text{spawn}}$ and $S^{(k)}_{\text{exit}}$ are chosen as the flux surfaces with *fixed* flux label s_{spawn} and s_{exit} for the field $\mathbf{B}^{(k)}$, and as $\mathbf{B}^{(k)}$ changes, the spawn and exit surfaces themselves must change. Using a toroidal angle ϕ and poloidal angle θ, the spawn surface $S^{(k)}_{\text{spawn}}$ can be described using with a function $\Gamma^{(k)}(\phi, \theta) = (x_0, y_0, z_0)$, where $\Gamma^{(k)} : [0, 1]^2 \to S^{(k)}_{\text{spawn}}$ is a one-to-one correspondence. Assuming the deuterium and tritium nuclei are at the same uniform temperature on $S^{(k)}_{\text{spawn}}$, alpha particles must be born uniformly on $S^{(k)}_{\text{spawn}}$. As a result, we model the uncertainty in the particle birth distribution using the pullback distribution $\nu^{(k)} = (\Gamma^{(k)})^{-1}_{\#}\text{Unif}(S^{(k)}_{\text{spawn}})$ on the fixed angle space $[0, 1]^2$.

As discussed above, to model the uncertainty in initial birth velocity, we follow the convention that particles are born isotropically in 3D velocity space, so that the initial parallel velocity v_0 has distribution $q := \text{Unif}(-V_{\max}, V_{\max})$ where $V_{\max}$ is the speed of a particle born with kinetic energy 3.5 MeV. The input

uncertainty we consider is then a random vector $\xi = (\phi, \theta, v_0)$ on the domain $D = [0,1]^2 \times [-V_{\max}, V_{\max}]$ whose distribution, at the kth outer-loop iteration, is given by the product distribution $\pi^{(k)} = \nu^{(k)} \times q$. For each outer-loop iteration k, we seek to estimate a metric of confinement $E_{\pi^{(k)}}[F^{(k)}]$ for the kth magnetic configuration, where $F^{(k)}(\xi)$ is a HFM describing the confinement of an energetic particle birthed with ξ. The confinement metric we consider in this work is the *mean modified lost time*, where

$$F^{(k)}(\xi) = F^{(k)}(\phi, \theta, v_0) = \min(\inf\{t : \mathbf{x}(t) \in S^{(k)}_{\text{exit}}, \mathbf{x}(0) = \Gamma^{(k)}(\phi, \theta), v(0) = v_0\}, T_{\max}) \tag{23}$$

where $\mathbf{x}(t) = (x(t), y(t), z(t))$ solves the dynamics (21-22) with magnetic field $\mathbf{B}^{(k)}$, initial position $\Gamma^{(k)}(\phi, \theta)$, and initial velocity v_0. The outer loop application we consider is stellarator optimization, and the outer-loop variables $\boldsymbol{\lambda}^{(k)}$ correspond to the Fourier coefficients of the stellarator coils and are the primary design optimization variables at hand. At each optimization iteration k, the $\boldsymbol{\lambda}^{(k)}$ specify the stellarator coils which then controls the magnetic field $\mathbf{B}^{(k)}$ through the Biot-Savart law.

This field $\mathbf{B}^{(k)}$ then dictates the spawn surface $S^{(k)}_{\text{spawn}}$, the exit surface $S^{(k)}_{\text{exit}}$, as well as the particle dynamics (21-22). All three of these determine the HFM $F^{(k)}$. The spawn surface $S^{(k)}_{\text{spawn}}$ specifies the one-to-one correspondence $\Gamma^{(k)}$ which then determines $\nu^{(k)}$ and thus $\pi^{(k)}$. A diagram summarizing the dependencies is displayed in Figure 2.

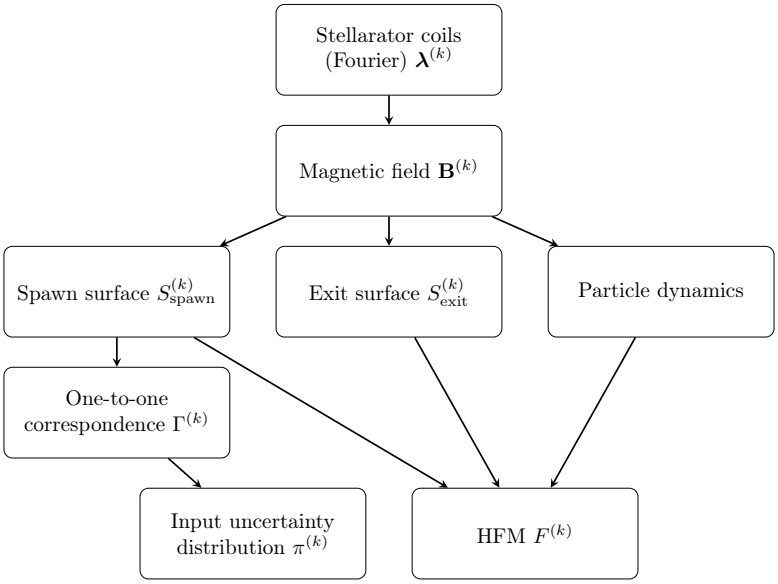

Figure 2: Flow chart demonstrating how outer-loop variables $\boldsymbol{\lambda}^{(k)}$ affect the input distribution $\pi^{(k)}$ and the HFM $F^{(k)}$ for the problem of energetic particle confinement.

5.2. Data-driven surrogate

We now discuss our choice of surrogate model G which we utilize in the four MF-based estimators (MF, ISMF, MFAIR, and ISMFAIR estimators). The primary reason we focus on constructing a data-driven surrogate models is that traditional sources of surrogate modeling are inappropriate for energetic particle motion [38]. For example, there is no hierarchy of simplified physics models to be leveraged [35] as the guiding center model already results from gyro-averaging the Lorentz force. Moreover, the multi-scale nature of alpha particle dynamics means that surrogates arising from coarser time steps are also unreliable. Additionally, popular projection based model reduction techniques such as proper orthogonal decomposition are largely ineffective since the problem is transport based [69].

For this purpose we utilize a data-driven surrogate model, namely an interpolant which we have previously done for energetic particle confinement [38]. In contrast to previous work which utilized piece-wise linear interpolation in 4 dimensions, here we utilize trigonometric interpolation in the angle variables (ϕ, θ) and piece-wise linear interpolation in the velocity variable v_0. We also leverage the domain knowledge that typically particles born with large parallel velocity $|v_0|$ are typically well confined, and thus we build our surrogate G on a truncated subdomain of $[-V_{\max}, V_{\max}]$. We note that this choice still *does not* take

advantage of inputs in the same manner that importance sampling does. Rather, this is for computationally efficiency, as we know particles outside the subdomain are more likely to be confined, and therefore we can have high surrogate accuracy in those regions.

More specifically, we select a subinterval $I_{\text{interpolate}} \subset [-V_{\max}, V_{\max}]$. Within this subinterval, we construct G to be an interpolant of $F^{(0)}$ using training-target pairs $\{(\xi_i, F^{(0)}(\xi_i)\}_{i=1}^{N_{\text{train}}}$ on the domain $[0,1]^2 \times I_{\text{interpolate}}$. For v_0 outside $I_{\text{interpolate}}$, we set G to be the constant value $T_{\max}$.

The choice of $I_{\text{interpolate}}$ is based on the discretion of the user and can be sourced from domain knowledge or pilot studies involving tracing a small number of particle trajectories. Further details on the interpolation used in our experiments is provided in Section 6.

Remark 5.1. *Since we take each $G^{(k)}$ to be a fixed model G which is trained at the initial optimization iteration, the performance of MF-based estimators along the optimization trajectory may depend greatly on the quality of the initial configuration with respect to the features of the HFM under consideration. Thus while MF with a data-driven surrogate alone may not provide enough variance reduction to lend estimation during the stellarator optimization loop tractable, it may be possible when using our meta estimators due to their quasi-multiplicative speedup.*

5.3. Data-driven biasing density

While there are multiple mechanisms for particle loss in stellarators, for many configurations certain classes of particles are significantly more at risk of escaping confinement than others. Namely, passing particles which rapidly ambulate the stellarator device are generally better confined than "trapped" particles which do not traverse the entire device but rather bounce near the edge slowly. These classes of orbits typically correspond to particles with larger $|v_0|$ (passing) and small $|v_0|$ (trapped).

In order to leverage this domain knowledge, we construct a biasing density of the form $\tilde{\pi}^{(k)} = \nu^{(k)} \times \tilde{q}$ where $\tilde{q}$ is designed to capture these features. More specifically, we employ a data-driven approach, using a Gaussian mixture model (GMM) with two components to construct $\tilde{q}$ using training data from the initial configuration.

$$\tilde{q}(v_0) = w_1 \mathcal{N}(m_1, \sigma_1^2) + w_2 \mathcal{N}(m_2, \sigma_2^2), \quad w_1 + w_2 = 1$$

While importance sampling can be powerful, it can be unstable if done incorrectly. Namely, since the IS estimator utilizes $F^{(k)}\pi^{(k)}/\tilde{\pi}^{(k)} = F^{(k)}q/\tilde{q}$, if the tails of $\tilde{q}$ are not sufficiently heavy, then $F^{(k)}q/\tilde{q}$ may not have finite variance. In practice, it is often desirable for $\tilde{q}$ to have significantly heavier tails than q to avoid numerical blowup in $q/\tilde{q}$, which is often why the optimal biasing density is not typically used even if known [49]. To ensure this does not happen for our data-driven biasing density, once our GMM is trained, we multiple all the GMM component variances by a safety factor.

Since our biasing density $\pi^{(k)}$ is the product distribution of $\nu^{(k)}$ for the spawn surface and $\tilde{q}$ for the parallel velocity, we see that our importance weights $\pi^{(k)}/\tilde{\pi}^{(k)} = q/\tilde{q}$ are independent of k. Since we are using the same surrogate model G at every optimization iteration, then the importance weighted surrogate $\tilde{G} = Gq/\tilde{q}$ is also independent of k. Further details of how we train our GMM for our numerical experiments are provided in Section 6.

Remark 5.2. *In this application we construct our data-driven biasing density from the initial configuration, just as for our surrogate model. However, in principle one could adapt the biasing density based on the outer-loop iteration k. We note that there has been work done in adapting the biasing density during the outer-loop, such as reusing samples from previous iterations to construct an optimal biasing density for the current iteration [53]. However, we leave the incorporation of adaptive biasing densities in meta estimators for future work.*

Remark 5.3. *Since the surrogate G only evaluates the interpolant within a subdomain of $[-V_{max}, V_{max}]$, G is significantly faster for particles which are spawned with v_0 outside this subdomain. As a result, we note that G is slower to evaluate under the biasing distribution $\tilde{\pi}^{(k)}$, since more particles will be born in this subdomain under $\tilde{\pi}^{(k)}$ compared to $\pi^{(k)}$ which means more interpolant evaluations will be needed as compared to sampling from $\pi^{(k)}$. This leads to a different cost ratio $\tilde{w}$ in the ISMF and ISMFAIR estimators compared to the cost ratio w in the MF and MFAIR estimators.*

6. Numerical Results

We tested our meta variance reduction estimators on two configurations. Firstly, on a configuration similar to the National Compact Stellarator Experiment (NCSX) [55] where there is no outer-loop present. This preliminary test is mainly designed to test the ISMF estimator when the configuration is constant. Next we tested our meta estimators on our primary scenario, an optimization trajectory of a recent quasi-axisymmetric (QA) stellarator configuration, henceforth referred to as LPQA2022 [18].

6.1. Numerical Setup

In this section, we consider the confinement of energetic particles for two different coil systems. The first coil system is the set of non-planar coils of the National Compact Stellarator Experiment (NCSX). NCSX is a compact high performance stellarator which was designed in the 1990s and early 2000s to have a magnetic field which approximates quasi-axisymmetry [70, 71]. It was partially built, but the project was canceled in 2008, because the estimated cost and schedule for completing the project grew as the technical requirements and risks became better understood [72]. The NCSX stellarator is composed of three unique modular coil shapes to which stellarator symmetry [73]and three-fold toroidal symmetry are applied [55]. The NCSX design also relies on planar toroidal field coils and poloidal field coils, which are not included in our study. As a result, our set of non-planar coils generates a magnetic field with significant departures from quasi-symmetry [55], and relatively poor confinement properties. The point of this numerical example is to study the validity and robustness of our meta estimators in situations in which the magnetic field is not well optimized, which can happen in the first few iterations of an optimization study.

The second coil system we consider is a coil system optimized to approximate the magnetic field with excellent quasi-symmetry and confinement properties recently found by Landreman and Paul [18]. It was recently shown that excellent approximations of this remarkable field could be generated by a realistic set of electromagnets [19]. The authors of this work obtained this coil system by constructing an optimization problem relying on the approach of the FOCUS coil design tool [74] and on several methods of the stellarator optimization code SIMSOPT [8]. The details of the optimization problem can be found at https://github.com/fredglaw/meta-multifidelity. From the point of view of the present article, this coil system is remarkable in two ways. First, the magnetic field it generates has excellent particle confinement. Second, we have access to the coil system at each iteration of the optimization process that was used to obtain the final high performance coils. Correspondingly, the purpose of this numerical example is two-fold. First, we are interested in the performance of our meta estimators for situations in which particles are very well confined, in contrast to our first numerical example. This example is therefore more relevant to the late stages of an stellarator optimization study. The second purpose of this numerical example is to investigate the capabilities of our meta estimators as it is applied at consecutive iterations of a design and optimization study.

In both coil systems we consider, the magnetic field is directly calculated from the coil geometry and coil currents, via the Biot-Savart law. For the coils, we rely on the common approximation that they are zero-thickness current-carrying filaments. Each filament is represented as a closed smooth curve in three dimensional space, which is described by a truncated Fourier series [74, 55, 56, 21]. The Fourier coefficients and the currents for the different coils for each example can be found at https://github.com/fredglaw/meta-multifidelity. Since the magnetic field is directly calculated from the Biot-Savart law, the existence of flux surfaces is not guaranteed, and not assumed. For the magnetic configurations we have considered however,

Configuration	N_ϕ	N_θ	N_v	$I_{\text{interpolate}}$
NCSX-like	40	40	40	$[-0.3V_{\max}, 0.3V_{\max}]$
LPQA2022	20	20	40	$[-0.25V_{\max}, 0.45V_{\max}]$

Table 1: Number of training points for interpolant G as well as the subdomain of training for the parallel velocity.

Configuration	m_1	m_2	σ_1^2	σ_2^2	w_1	w_2
NCSX-like	4.084×10^{-2}	-3.632×10^{-2}	1.147×10^{-2}	8.048×10^{-3}	0.4994	0.5006
LPQA2022	0.1406	0.2024	2.107×10^{-3}	3.376×10^{-3}	0.4846	0.5154

Table 2: Means, variances, and weights for trained GMM biasing densities using 5000 training-target pairs. Training data is scaled to lie in $[-1, 1]$.

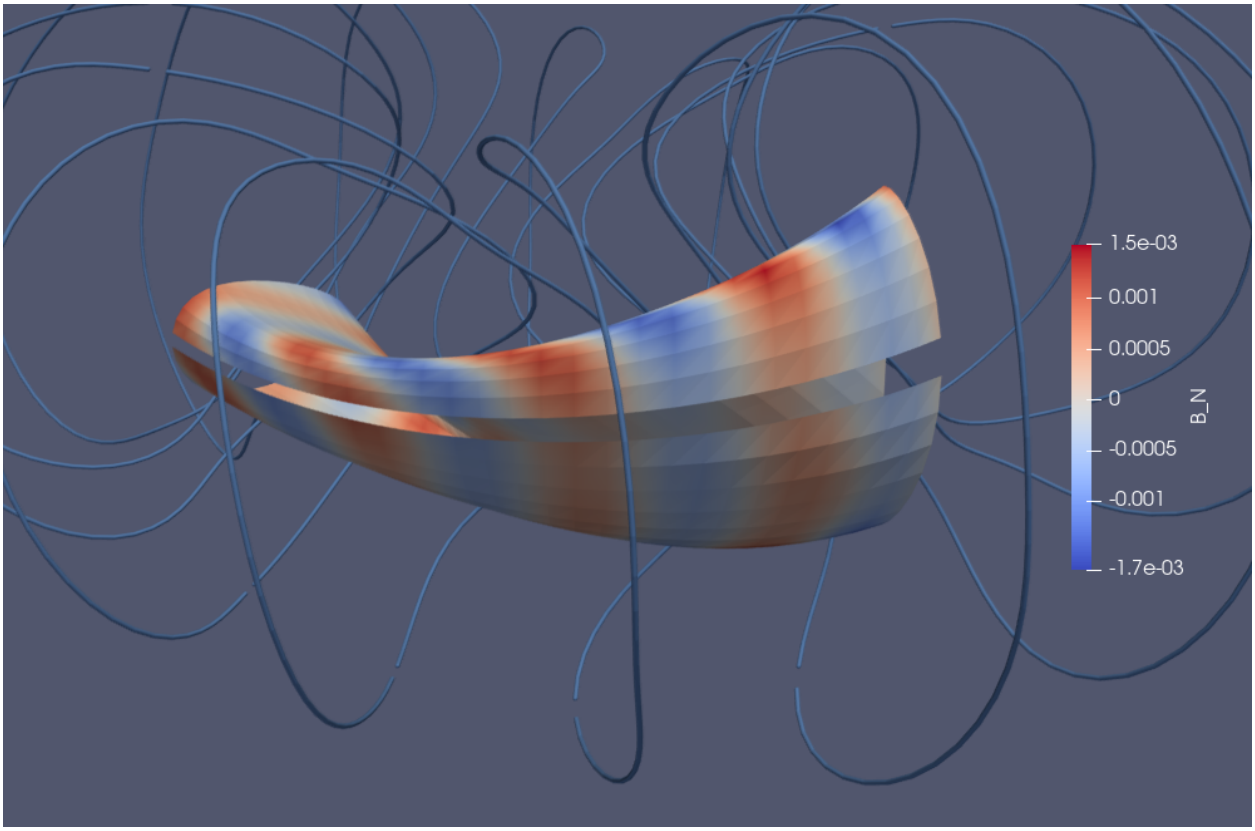

Figure 3: Half field period of $S^{(1000)}_{\text{exit}}$ from the coil optimization with LPQA2022 as the target configuration. We see the discrete coil effect through the striped normal component of the $\mathbf{B}$ field on the surface.

we were able to identify flux surfaces, and for fixed flux labels s_{spawn} and s_{exit}, we numerically construct $S^{(k)}_{\text{spawn}}$ and $S^{(k)}_{\text{exit}}$ with the method described in [20]. For the NCSX-like coil set we use $s_{\text{spawn}} = 0.005$ and $s_{\text{exit}} = 0.3$, and for LPQA2022 we use $s_{\text{spawn}} = 0.1$ and $s_{\text{exit}} = 0.3$.

Particles are birthed in the angle space $[0, 1]^2$ according to the pullback distribution $\nu^{(k)}$ discussed in Section 5. Samples from $\nu^{(k)}$ are drawn by rejection sampling using a uniform proposal, where the target distribution is proportional to the Jacobian of $\Gamma^{(k)}$. The normalizing constant for this target density is computed to sufficiently high accuracy using trapezoidal quadrature for the double periodic $\Gamma^{(k)}$. Particle parallel velocity is sampled either uniformly according to q or according to the GMM $\tilde{q}$.

Given initial conditions, particles are traced by solving the guiding-center equations (21-22) using an adaptive ODE integrator, namely Dormand-Prince, and followed until either time $T_{\text{max}} = 10^{-3}$ s or until they reach the exit surface $S^{(k)}_{\text{exit}}$ and are classified as lost. Classification of loss is done by tracking particles with a signed distance function which is positive in the volume contained by the exit surface $S^{(k)}_{\text{exit}}$, zero on the exit surface, and negative outside the exit surface.

We note that when integrating equations (21-22), we use a polynomial interpolant of the Biot-Savart magnetic field which is precomputed on a mesh in cylindrical coordinates. The interpolated magnetic field will not be exactly divergence free, however we ensure that the interpolation error is sufficiently small compared to the error of the numerical integration.

For all our numerical tests, we use protons with 9 keV of kinetic energy as proxies for alpha particles with 3.5 MeV in a reactor scale device with the dimensions of the ARIES-CS stellarator power plant design [75, 76, 56]. Specifically, if we call $\rho_\star$ the ratio of the small radius of the helical motion of particles around field lines to the characteristic radius of the plasma cross section, then with a kinetic energy of 9 keV, the proton trajectories we compute have the same $\rho_\star$ in the magnetic configurations we consider in this manuscript as 3.5 MeV alpha particles in the ARIES-CS device. Since $\rho_\star$ is the key non-dimensional parameter of interest for particle transport, we thus expect our results to be relevant for fusion power plant devices. As a consequence, physical parameters in our numerical experiments, such as maximum velocity, mass, charge,

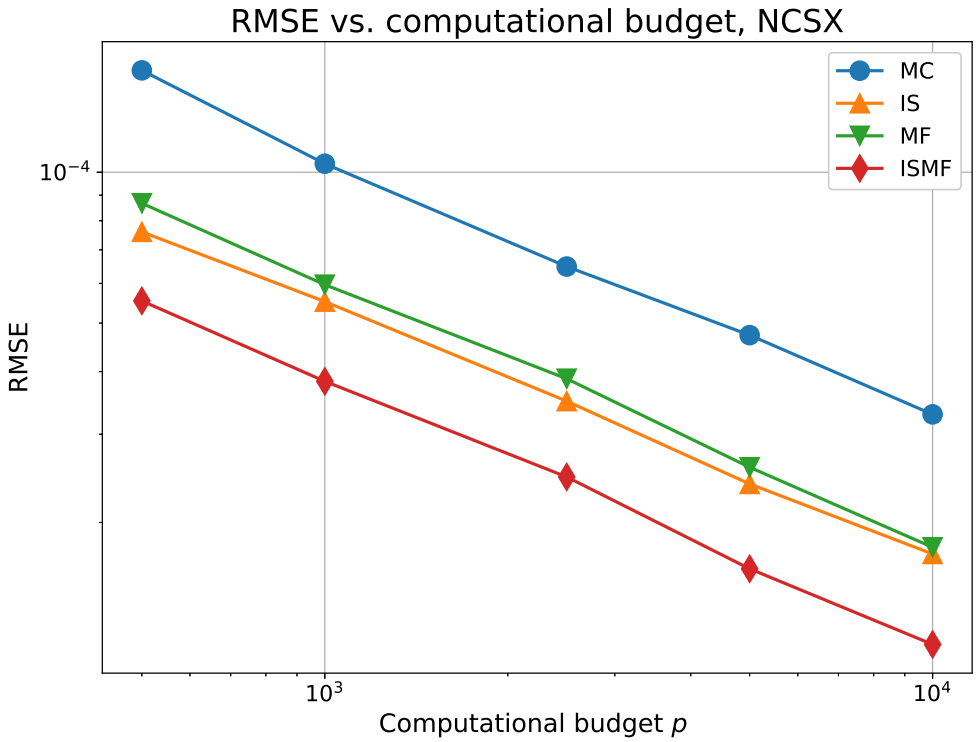

Figure 4: RMSE versus computational budget p for the NCSX-like configuration. While both IS and MF provide variance reduction on their own, the combined MFIS estimator improves on both.

and field strength, are scaled for protons.

For our data-driven surrogate, we utilize trigonometric interpolation in the angle variables (ϕ, θ) and piece-wise linear interpolation in the velocity variable v_0 on the domain $[0,1] \times [0,1] \times I_{\text{interpolate}}$. This is done using N_ϕ, N_θ, N_v equispaced points for ϕ, θ, v_0 respectively. The number of grid points and parallel velocity subdomain used to train the surrogate in each coil set case is provided in Table 1. Each subdomain was chosen heuristically by examining a pilot study of 1000 particle trajectories. While more systematic choices of $I_{\text{interpolate}}$ are certainly possible, we utilize this ad-hoc choice since the focus in this work is on the efficacy of meta estimators provided G and not the construction of G itself.

For our GMM $\tilde{q}$, we generate training data by drawing samples of v_0 from a centered Gaussian with most of its mass concentrated near 0. We choose this initial Gaussian to be intentionally heavy-tailed as we simply want to generate training data that is reflective of the domain knowledge we are leveraging. We note that this is similar to the subinterval $I_{\text{interpolate}}$ utilized our surrogate model, as both are drawing from the same source of domain knowledge. The distinction however is that for the GMM we simply use this as a guide to generate the training data itself. In this application we choose the initial Gaussian to place 95% of its mass on the interval $[-0.3V_{\max}, 0.3V_{\max}]$. We train our Gaussian using the expectation maximization algorithm in Scikit-learn [77] using 5000 training-target pairs where the training data is drawn i.i.d. from the product distribution of $\nu^{(k)}$ and this initial Gaussian, and the targets are the classification of if a particle is lost or not. The training pairs are scaled by $V_{\max}$ to lie in $[-1, 1]$. To avoid numerical blowup of our trained $\tilde{q}$, we multiple the standard deviation of each component by a safety factor of 2.25. This training procedure is the same for the NCSX and LPQA2022 case, using training data from those cases respectively.

Further details on the coil parameterizations and currents, the Biot-Savart evaluations, the particle trajectory integrator, and the classifier for the numerical examples in this article were implemented in the SIMSOPT package [8], available at https://github.com/hiddenSymmetries/simsopt.

6.2. NCSX-like configuration

We first compare the ISMF estimator against the MC, MF, and IS estimators on the magnetic configuration arising from the NCSX coil set. In this case there is no outer-loop application, and thus we shall omit the notational dependence on k for this case. Using 1000 samples, we report a correlation of $\rho_\pi(F, G) \approx 0.883$ and $\rho_{\tilde{\pi}}(\widetilde{F}, \widetilde{G}) \approx 0.788$. The cost ratios w and $\tilde{w}$ are reported in Table 3 where we see that the cost ratio is 7 times smaller when sampling from $\tilde{\pi}$ compared to sampling from π. This is due to our trained GMM $\tilde{q}$ having a bulk of its probability mass concentrated in $I_{\text{interpolate}}$ as can be seen in Table 2. As a result, when sampling the parallel velocity from $\tilde{q}$, more samples land in $I_{\text{interpolate}}$ leading to interpolant evaluation which is more costly than simply assigning $T_{\max}$ as the output.

We note that when switching from using F, G in the MF estimator to $\widetilde{F}, \widetilde{G}$ in the ISMF estimator, both the correlation and cost ratios *decrease* which suggests that $c_{\text{MF}}(\widetilde{F}, \widetilde{G}, \tilde{\pi}) > c_{\text{MF}}(F, G, \pi)$. However, provided that that the variance reduction c_{IS} from the IS estimator is sufficiently small, we may expect the ISMF estimator to outperform the MF and IS estimators alone.

Using computational budgets $p = 500, 1000, 2500, 5000, 10000$ we generated 200 replicates of the MC, IS, MF, and ISMF estimators. In Figure 4 we plot the RMSE of each estimator as a function of the budget p for the scaled mean modified loss time $E_\pi[F]/T_{\max}$. We see that while both the MF and IS estimators provide variance reduction compared to the MC estimator, the combined ISMF estimator outperforms the constituent MF and IS estimators. Based on empirically measured c_{IS}, the correlations $\rho_\pi(F, G)$ and $\rho_{\tilde{\pi}}(\widetilde{F}, \widetilde{G})$, and the cost ratios in Table 3, our numerical results agree with theoretical estimates to leading order. Thus we indeed observe quasi-multiplicative speedup for our ISMF estimator.

6.3. LPQA2022 configuration

Next we test all our meta estimators on a coil optimization trajectory whose target configuration is the LPQA2022 configuration. Since this optimization trajectory starts with circular coils, we focus on estimating particle confinement in the second half of the trajectory. That is, we train our surrogate model G and GMM $\tilde{q}$ starting at the $k = 500$th optimization trajectory and test our meta estimators at the $k = 600, 700, 800, 900, 1000$ optimization iterations.

In Table 3 we see a similar trend in cost ratios when introducing the GMM $\tilde{q}$ for sampling the parallel velocity as in the NCSX case. The reasoning is the same as the NCSX case, the trained $\tilde{q}$ places much of its mass on $I_{\text{interpolate}}$ which makes the surrogate model more expensive on average. We note that $I_{\text{interpolate}}$ is not centered for the LPQA2022 case, see Table 1. This was observed in our pilot study which is why we chose a non-centered interval. Although the data used to train our GMM was drawn from a centered Gaussian, we see the trained $\tilde{q}$ is both shifted and skewed to the right, see Table 2. Thus this overlap between the $I_{\text{interpolate}}$ for the surrogate and where $\tilde{q}$ places most of its mass was not entirely chosen by us. Rather, we decided $I_{\text{interpolate}}$ and the expectation-maximization trained $\tilde{q}$ to place mass there as well.

Using a budget of $p = 5000$, we generated 250 replicates of MC, MF, IS, AIR, MFAIR, ISMF, ISAIR, and ISMFAIR estimators. In Figure 5 we plot the RMSE for each estimator for the scaled mean modified loss time $E_\pi^{(k)}[F^{(k)}]/T_{\max}$ as a function of the optimization iteration k. Recall that although our surrogate and GMM

Configuration	w	$\tilde{w}$
NCSX-like	355	54
LPQA2022	242	35

Table 3: Approximate cost ratios w and $\tilde{w}$ for the NCSX-like and LPQA2022 configurations. Estimated using 5000 samples at the initial configuration.

k	$\text{Var}_{\pi^{(k)}}(F^{(k)})$	$\text{Var}_{\pi^{(k)}}(F^{(k-1)}W_k)$	$\text{Var}_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)})$	$\text{Var}_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k-1)}W_k)$	$c_{\text{IS}}^{(k)}$
600	2.643e-02	2.643e-02	2.769e-03	2.769e-03	0.1047
700	2.712e-02	2.710e-02	2.761e-03	2.762e-03	0.1018
800	2.686e-02	2.686e-02	2.769e-03	2.768e-03	0.1031
900	2.684e-02	2.682e-02	2.766e-03	2.766e-03	0.1030
1000	2.681e-02	2.682e-02	2.779e-03	2.779e-03	0.1037

Table 4: Measured variances and estimates of variance reduction using importance sampling for different k in the LPQA2022 case. Each variance is the average of 250 replicates of sample variances, where each sample variance uses $p/2 = 2500$ samples. Each $c_{\text{IS}}^{(k)}$ is estimated as $\text{Var}_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)})/\text{Var}_{\pi^{(k)}}(F^{(k)})$ using the measured variances in the table.

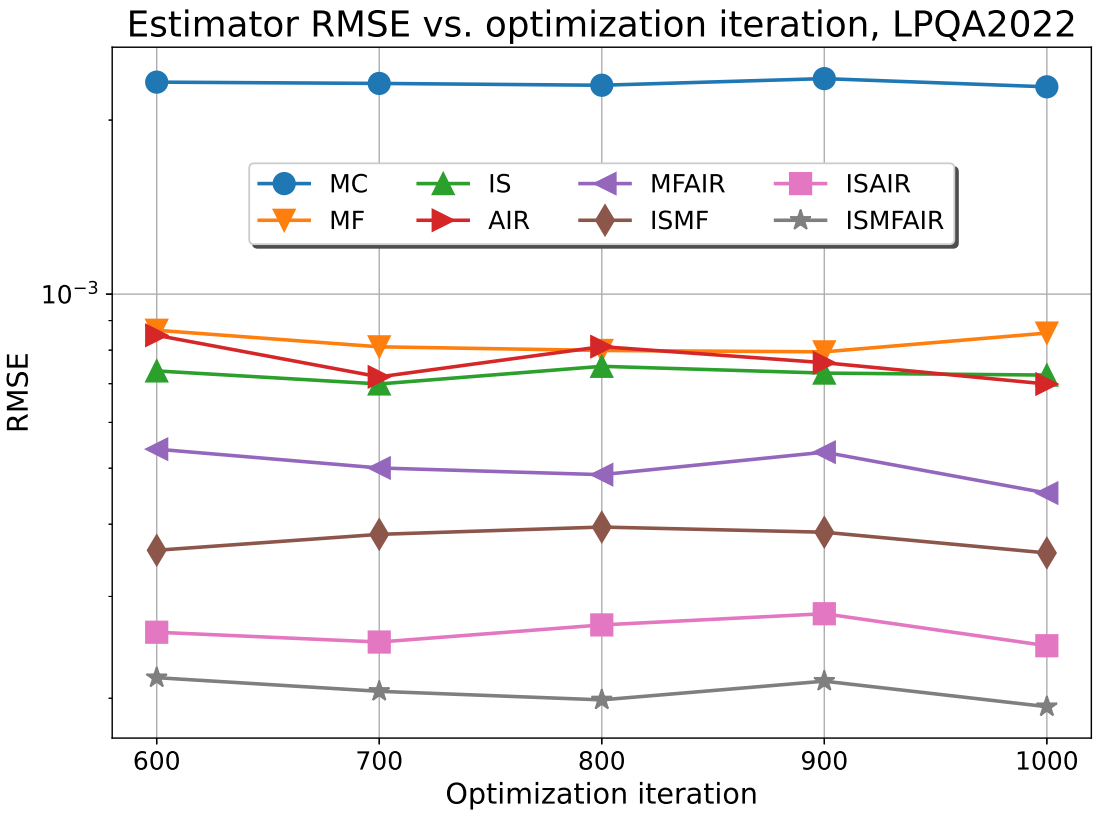

Figure 5: Measured RMSE of each estimator at $k = 600, 700, 800, 900, 1000$. Each variance is estimated using 250 estimator replicates, with each estimator using a budget of $p = 5000$ HFM evaluations.

were trained only using data at the $k = 500$ iteration, the variance reduction for all estimators is practically constant for $k = 600, 700, 800, 900, 1000$. This suggests that for late-stage optimization applications, data-driven surrogates and biasing densities may be highly effective even when not adapted at each iteration. Moreover, we see that each meta estimator, i.e. MFAIR, ISMF, ISAIR, and ISMFAIR, outperforms its constituent estimators.

In Figure 6 we plot the speedup of each constituent and meta estimator compared to the MC estimator, averaged over the values at $k = 600, 700, 800, 900, 1000$. We also plot the theoretical speedup for the MF, IS, ISMF estimators as well as the asympotic speedup for the AIR, MFAIR, ISAIR, and ISMFAIR estimators. The theoretical and asymptotic values are estimated using the variances and correlations in Tables 4 and 5 respectively, and then averaged over $k = 600, 700, 800, 900, 1000$. We see excellent agreement between the measured speedups and the theoretical and asymptotic speedups for all of our estimators. For the ISMFAIR estimator which leverages all three of our constituent estimators, we measure over two orders of magnitude speedup compared to Monte Carlo.

Recall that in deriving our asymptotic variance reduction formulas, we relied on certain assumptions about the convergence of variances and correlations. In deriving the asymptotic variance reduction for the AIR estimator, we assumed that $\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k)$ converged and that $\mathrm{Var}_{\pi^{(k-1)}}(F^{(k-1)})/\mathrm{Var}_{\pi^{(k)}}(F^{(k-1)}W_k) \to 1$ as $k \to \infty$. Examining Tables 4 and 5, we approximately verify that this is indeed the case. Note that while we do not measure $\mathrm{Var}_{\pi^{(k-1)}}(F^{(k-1)})$, we observe that $\mathrm{Var}_{\pi^{(k)}}(F^{(k)})$ remains primarily constant in k, and thus deduce that $\mathrm{Var}_{\pi^{(k)}}(F^{(k)}) \approx \mathrm{Var}_{\pi^{(k-1)}}(F^{(k-1)})$ for large k. Similar assumptions hold for the asymptotic variance reduction of the ISAIR estimator with $\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)}, \widetilde{F}^{(k-1)}W_k)$ converging and with $\mathrm{Var}_{\tilde{\pi}^{(k-1)}}(\widetilde{F}^{(k-1)})/\mathrm{Var}_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k-1)}W_k) \to 1$ as $k\infty$. In deriving the asymptotic variance reduction for MFAIR, we assumed that $\rho_{\pi^{(k)}}(F^{(k)}, G)$ and $\rho_{\pi^{(k)}}(F^{(k-1)}W_k, G)$ both converged to the same

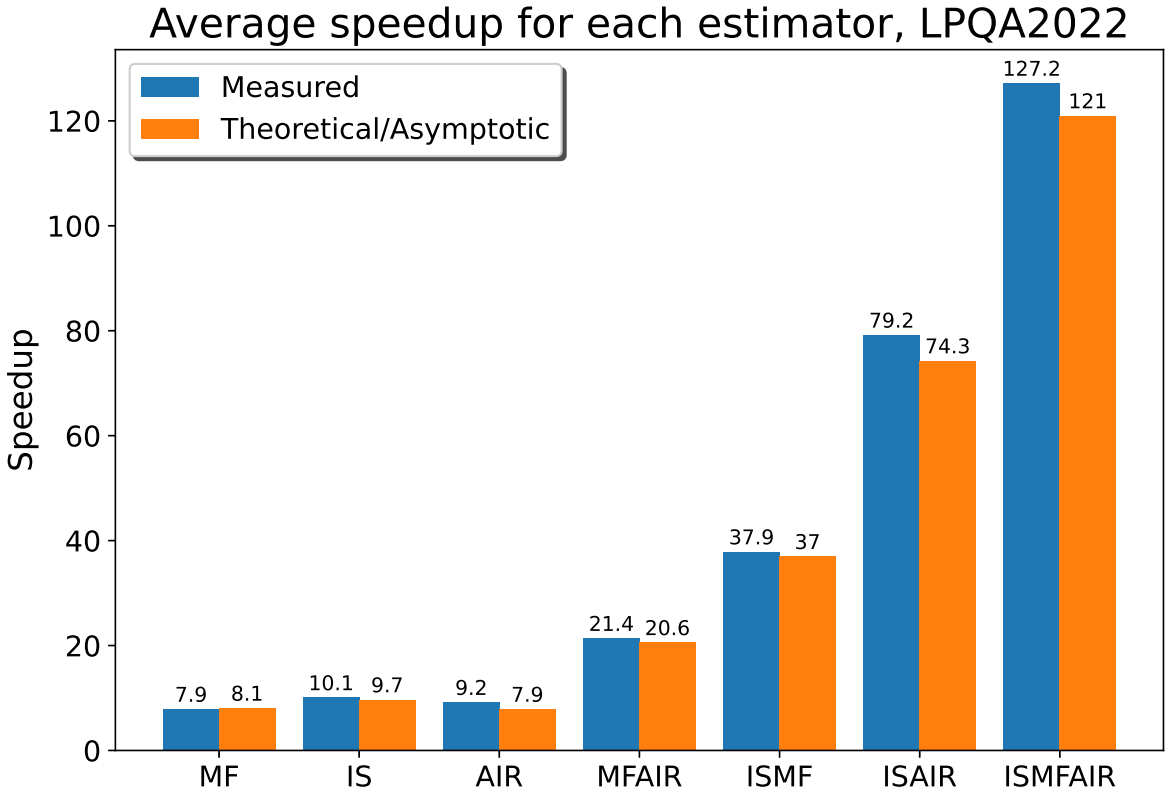

Figure 6: Measured and theoretical speedups for each method compared to the MC estimator with equivalent cost, averaged over their values at each $k = 600, 700, 800, 900, 1000$. Each speedup is measured using 250 estimator replicates, where each estimator used a budget of $p = 5000$ HFM evaluations. For each IR-based method, the approximation of asymptotic variance reduction is presented. Theoretical and asymptotic estimates at each k are computed using $c_{\text{IS}}^{(k)}$ from Table 4 and measured correlations in Table 5.

value as $k \to \infty$. Moreover, we assumed that $\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k)$ converged as $k \to \infty$ and that $\text{Var}_{\pi^{(k-1)}}(F^{(k-1)})/\text{Var}_{\pi^{(k)}}(F^{(k-1)}W_k) \to 1$ as $k \to \infty$. Examining the measured values in Tables 4 and 5, we can again approximately verify these assumptions. Likewise for the ISMFAIR estimator, we observe that $\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)}, G)$ and $\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k-1)}W_k, G)$ roughly converge to the same value, $\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)}, \widetilde{F}^{(k-1)}W_k)$ seems to converge, and $\text{Var}_{\pi^{(k-1)}}(F^{(k-1)})/\text{Var}_{\pi^{(k)}}(F^{(k-1)}W_k) \to 1$ as $k \to \infty$ is also approximately verified.

References

[1] A. Sagara, Y. Igitkhanov, F. Najmabadi, Review of stellarator/heliotron design issues towards mfe demo, Fusion Engineering and Design 85 (7) (2010) 1336–1341, proceedings of the Ninth International Symposium on Fusion Nuclear Technology. doi:https://doi.org/10.1016/j.fusengdes.2010.03.041.
URL https://www.sciencedirect.com/science/article/pii/S0920379610001122

[2] P. Helander, C. Beidler, T. Bird, M. Drevlak, Y. Feng, R. Hatzky, F. Jenko, R. Kleiber, J. Proll, Y. Turkin, et al., Stellarator and tokamak plasmas: a comparison, Plasma Physics and Controlled Fusion 54 (12) (2012) 124009.

[3] A. H. Boozer, Stellarators as a fast path to fusion, Nuclear Fusion 61 (9) (2021) 096024. doi:10.1088/1741-4326/ac170f.
URL https://doi.org/10.1088/1741-4326/ac170f

k	$\rho_{\pi^{(k)}}(F^{(k)}, G)$	$\rho_{\pi^{(k)}}(F^{(k-1)}W_k, G)$	$\rho_{\pi^{(k)}}(F^{(k)}, F^{(k-1)}W_k)$
600	0.9575	0.9575	0.9980
700	0.9581	0.9583	0.9975
800	0.9571	0.9567	0.9980
900	0.9553	0.9551	0.9977
1000	0.9571	0.9573	0.9986
	$\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)}, \widetilde{G})$	$\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k-1)}W_k, \widetilde{G})$	$\rho_{\tilde{\pi}^{(k)}}(\widetilde{F}^{(k)}, \widetilde{F}^{(k-1)}W_k)$
600	0.9371	0.9371	0.9977
700	0.9352	0.9355	0.9972
800	0.9348	0.9347	0.9976
900	0.9350	0.9350	0.9974
1000	0.9348	0.9348	0.9987

Table 5: Measured correlations between different models under different input distributions in the LPQA2022 case. Each correlation is estimated by averaging over 250 replicates of sample correlation. Sample correlations between high-fidelity were estimated using $p/2 = 2500$ samples, whereas sample correlations between high-fidelity and surrogate models were estimated using n samples, where n is the number of HFM evaluations dictated by the MF estimator with relevant models.

[4] S. P. Hirshman, D. A. Spong, J. C. Whitson, V. E. Lynch, D. B. Batchelor, B. A. Carreras, J. A. Rome, Transport optimization and mhd stability of a small aspect ratio toroidal hybrid stellarator, Phys. Rev. Lett. 80 (1998) 528–531. doi:10.1103/PhysRevLett.80.528.
URL https://link.aps.org/doi/10.1103/PhysRevLett.80.528

[5] D. A. Spong, S. P. Hirshman, J. C. Whitson, D. B. Batchelor, B. A. Carreras, V. E. Lynch, J. A. Rome, J* optimization of small aspect ratio stellarator/tokamak hybrid devices, Physics of Plasmas 5 (5) (1998) 1752–1758. arXiv:https://doi.org/10.1063/1.872844, doi:10.1063/1.872844.
URL https://doi.org/10.1063/1.872844

[6] M. Drevlak, C. Beidler, J. Geiger, P. Helander, Y. Turkin, Optimisation of stellarator equilibria with ROSE, Nuclear Fusion 59 (1) (2018) 016010. doi:10.1088/1741-4326/aaed50.
URL https://doi.org/10.1088/1741-4326/aaed50

[7] S. Lazerson, C. Caoxiang Zhu, J. Schmitt, et al., STELLOPT (2021).
URL https://github.com/PrincetonUniversity/STELLOPT

[8] M. Landreman, B. Medasani, F. Wechsung, A. Giuliani, R. Jorge, C. Zhu, Simsopt: A flexible framework for stellarator optimization, Journal of Open Source Software 6 (65) (2021) 3525. doi:10.21105/joss.03525.
URL https://doi.org/10.21105/joss.03525

[9] R. C. Wolf, A. Alonso, S. Äkäslompolo, J. Baldzuhn, M. Beurskens, C. D. Beidler, C. Biedermann, H.-S. Bosch, S. Bozhenkov, R. Brakel, H. Braune, S. Brezinsek, K.-J. Brunner, H. Damm, A. Dinklage, P. Drewelow, F. Effenberg, Y. Feng, O. Ford, G. Fuchert, Y. Gao, J. Geiger, O. Grulke, N. Harder, D. Hartmann, P. Helander, B. Heinemann, M. Hirsch, U. Höfel, C. Hopf, K. Ida, M. Isobe, M. W. Jakubowski, Y. O. Kazakov, C. Killer, T. Klinger, J. Knauer, R. König, M. Krychowiak, A. Langenberg, H. P. Laqua, S. Lazerson, P. McNeely, S. Marsen, N. Marushchenko, R. Nocentini, K. Ogawa, G. Orozco, M. Osakabe, M. Otte, N. Pablant, E. Pasch, A. Pavone, M. Porkolab, A. Puig Sitjes, K. Rahbarnia, R. Riedl, N. Rust, E. Scott, J. Schilling, R. Schroeder, T. Stange, A. von Stechow, E. Strumberger, T. Sunn Pedersen, J. Svensson, H. Thomson, Y. Turkin, L. Vano, T. Wauters, G. Wurden, M. Yoshinuma, M. Zanini, D. Zhang, Performance of wendelstein 7-x stellarator plasmas during the first divertor operation phase, Physics of Plasmas 26 (8) (2019) 082504. arXiv:https://doi.org/10.1063/1.5098761, doi:10.1063/1.5098761.
URL https://doi.org/10.1063/1.5098761

[10] J. Alonso, I. Calvo, D. Carralero, J. Velasco, J. García-Regaña, I. Palermo, D. Rapisarda, Physics design point of high-field stellarator reactors, Nuclear Fusion 62 (3) (2022) 036024. doi:10.1088/1741-4326/

ac49ac.
URL https://doi.org/10.1088/1741-4326/ac49ac

[11] E. Paul, A. Bhattacharjee, M. Landreman, D. Alex, J. Velasco, R. Nies, Energetic particle loss mechanisms in reactor-scale equilibria close to quasisymmetry, arXiv preprint arXiv:2208.02351 (2022).

[12] L.-P. Ku, P. Garabedian, J. Lyon, A. Turnbull, A. Grossman, T. Mau, M. Zarnstorff, A. Team, Physics design for aries-cs, Fusion Science and Technology 54 (3) (2008) 673–693.

[13] J. Velasco, I. Calvo, S. Mulas, E. Sánchez, F. Parra, Á. Cappa, the W7-X Team, A model for the fast evaluation of prompt losses of energetic ions in stellarators, Nuclear Fusion 61 (11) (2021) 116059. doi:10.1088/1741-4326/ac2994.
URL https://doi.org/10.1088/1741-4326/ac2994

[14] W. Lotz, P. Merkel, J. Nuhrenberg, E. Strumberger, Collisionless alpha -particle confinement in stellarators, Plasma Physics and Controlled Fusion 34 (6) (1992) 1037–1052. doi:10.1088/0741-3335/34/6/010.
URL https://doi.org/10.1088/0741-3335/34/6/010

[15] A. Subbotin, M. Mikhailov, V. Shafranov, M. Isaev, C. Nührenberg, J. Nührenberg, R. Zille, V. Nemov, S. Kasilov, V. Kalyuzhnyj, W. Cooper, Integrated physics optimization of a quasi-isodynamic stellarator with poloidally closed contours of the magnetic field strength, Nuclear Fusion 46 (11) (2006) 921–927. doi:10.1088/0029-5515/46/11/006.
URL https://doi.org/10.1088/0029-5515/46/11/006

[16] S. Henneberg, M. Drevlak, C. Nührenberg, C. D. Beidler, Y. Turkin, J. Loizu, P. Helander, Properties of a new quasi-axisymmetric configuration, Nuclear Fusion 59 (2) (2019) 026014.

[17] C. G. Albert, S. V. Kasilov, W. Kernbichler, Accelerated methods for direct computation of fusion alpha particle losses within, stellarator optimization, Journal of Plasma Physics 86 (2) (2020) 815860201. doi:10.1017/S0022377820000203.

[18] M. Landreman, E. Paul, Magnetic fields with precise quasisymmetry for plasma confinement, Phys. Rev. Lett. 128 (2022) 035001. doi:10.1103/PhysRevLett.128.035001.
URL https://link.aps.org/doi/10.1103/PhysRevLett.128.035001

[19] F. Wechsung, M. Landreman, A. Giuliani, A. Cerfon, G. Stadler, Precise stellarator quasi-symmetry can be achieved with electromagnetic coils, Proceedings of the National Academy of Sciences 119 (13) (2022) e2202084119.

[20] A. Giuliani, F. Wechsung, G. Stadler, A. Cerfon, M. Landreman, Direct computation of magnetic surfaces in boozer coordinates and coil optimization for quasisymmetry, Journal of Plasma Physics 88 (4) (2022) 905880401. doi:10.1017/S0022377822000563.

[21] F. Wechsung, A. Giuliani, M. Landreman, A. Cerfon, G. Stadler, Single-stage gradient-based stellarator coil design: stochastic optimization, Nuclear Fusion 62 (7) (2022) 076034. doi:10.1088/1741-4326/ac45f3.
URL https://doi.org/10.1088/1741-4326/ac45f3

[22] V. V. Nemov, S. V. Kasilov, W. Kernbichler, G. O. Leitold, Poloidal motion of trapped particle orbits in real-space coordinates, Physics of Plasmas 15 (5) (2008) 052501. arXiv:https://doi.org/10.1063/1.2912456, doi:10.1063/1.2912456.
URL https://doi.org/10.1063/1.2912456

[23] A. Bader, M. Drevlak, D. T. Anderson, B. J. Faber, C. C. Hegna, K. M. Likin, J. C. Schmitt, J. N. Talmadge, Stellarator equilibria with reactor relevant energetic particle losses, Journal of Plasma Physics 85 (5) (2019) 905850508. doi:10.1017/S0022377819000680.

[24] A. Bader, D. Anderson, M. Drevlak, B. Faber, C. Hegna, S. Henneberg, M. Landreman, J. Schmitt, Y. Suzuki, A. Ware, Modeling of energetic particle transport in optimized stellarators, Nuclear Fusion 61 (11) (2021) 116060. doi:10.1088/1741-4326/ac2991.
URL https://doi.org/10.1088/1741-4326/ac2991

[25] P. Helander, Theory of plasma confinement in non-axisymmetric magnetic fields, Reports on Progress in Physics 77 (8) (2014) 087001. doi:10.1088/0034-4885/77/8/087001.
URL https://doi.org/10.1088/0034-4885/77/8/087001

[26] M. Landreman, S. Buller, M. Drevlak, Optimization of quasisymmetric stellarators with self-consistent bootstrap current and energetic particle confinement, arXiv preprint arXiv:2205.02914 (2022).

[27] F. I. Parra, I. Calvo, P. Helander, M. Landreman, Less constrained omnigeneous stellarators, Nuclear Fusion 55 (3) (2015) 033005.

[28] G. G. Plunk, M. Landreman, P. Helander, Direct construction of optimized stellarator shapes. part 3. omnigenity near the magnetic axis, Journal of Plasma Physics 85 (6) (2019) 905850602. doi:10.1017/S002237781900062X.

[29] R. Jorge, G. Plunk, M. Drevlak, M. Landreman, J.-F. Lobsien, K. C. Mata, P. Helander, A single-field-period quasi-isodynamic stellarator, arXiv preprint arXiv:2205.05797 (2022).

[30] H. Greuner, B. Böswirth, J. Boscary, G. Hofmann, B. Mendelevitch, H. Renner, R. Rieck, Final design of w7-x divertor plasma facing components—tests and thermo-mechanical analysis of baffle prototypes, Fusion Engineering and Design 66-68 (2003) 447–452, 22nd Symposium on Fusion Technology. doi:https://doi.org/10.1016/S0920-3796(03)00193-5.
URL https://www.sciencedirect.com/science/article/pii/S0920379603001935

[31] B. Sorbom, J. Ball, T. Palmer, F. Mangiarotti, J. Sierchio, P. Bonoli, C. Kasten, D. Sutherland, H. Barnard, C. Haakonsen, et al., Arc: A compact, high-field, fusion nuclear science facility and demonstration power plant with demountable magnets, Fusion Engineering and Design 100 (2015) 378–405.

[32] G. Bongiovì, G. Marra, R. Mozzillo, A. Tarallo, Heterogeneous design and mechanical analysis of helias 5-b helium-cooled pebble bed breeding blanket concept, International Journal of Energy Research 46 (3) (2022) 2748–2770. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/er.7343, doi:https://doi.org/10.1002/er.7343.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/er.7343

[33] L. W. T. Ng, K. E. Willcox, Multifidelity approaches for optimization under uncertainty, International Journal for Numerical Methods in Engineering 100 (10) (2014) 746–772. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.4761, doi:https://doi.org/10.1002/nme.4761.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.4761

[34] B. Peherstorfer, K. Willcox, M. Gunzburger, Optimal model management for multifidelity monte carlo estimation, SIAM Journal on Scientific Computing 38 (5) (2016) A3163–A3194.

[35] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, SIAM Review 60 (3) (2018) 550–591. arXiv:https://doi.org/10.1137/16M1082469, doi:10.1137/16M1082469.
URL https://doi.org/10.1137/16M1082469

[36] B. Peherstorfer, Multifidelity Monte Carlo estimation with adaptive low-fidelity models, SIAM/ASA Journal on Uncertainty Quantification 7 (2019) 579–603.

[37] J. Konrad, I.-G. Farcaş, B. Peherstorfer, A. Di Siena, F. Jenko, T. Neckel, H.-J. Bungartz, Data-driven low-fidelity models for multi-fidelity monte carlo sampling in plasma micro-turbulence analysis, Journal of Computational Physics 451 (2022) 110898. doi:https://doi.org/10.1016/j.jcp.2021.110898.
URL https://www.sciencedirect.com/science/article/pii/S0021999121007932

[38] F. Law, A. Cerfon, B. Peherstorfer, Accelerating the estimation of collisionless energetic particle confinement statistics in stellarators using multifidelity monte carlo, Nuclear Fusion 62 (7) (2022) 076019. doi:10.1088/1741-4326/ac4777.
URL https://doi.org/10.1088/1741-4326/ac4777

[39] M. B. Giles, Multilevel monte carlo path simulation, Operations Research 56 (3) (2008) 607–617. arXiv:https://doi.org/10.1287/opre.1070.0496, doi:10.1287/opre.1070.0496.
URL https://doi.org/10.1287/opre.1070.0496

[40] K. A. Cliffe, M. B. Giles, R. Scheichl, A. L. Teckentrup, Multilevel monte carlo methods and applications to elliptic pdes with random coefficients, Computing and Visualization in Science 14 (1) (2011) 3. doi:10.1007/s00791-011-0160-x.
URL https://doi.org/10.1007/s00791-011-0160-x

[41] A.-L. Haji-Ali, F. Nobile, R. Tempone, Multi-index monte carlo: when sparsity meets sampling, Numerische Mathematik 132 (4) (2016) 767–806. doi:10.1007/s00211-015-0734-5.
URL https://doi.org/10.1007/s00211-015-0734-5

[42] G. Dimarco, L. Pareschi, Multi-scale control variate methods for uncertainty quantification in kinetic equations, Journal of Computational Physics 388 (2019) 63 – 89.

[43] G. Dimarco, L. Pareschi, Multiscale variance reduction methods based on multiple control variates for kinetic equations with uncertainties, Multiscale Modeling & Simulation 18 (1) (2020) 351–382.

[44] I.-G. Farcas, T. Görler, H.-J. Bungartz, F. Jenko, T. Neckel, Sensitivity-driven adaptive sparse stochastic approximations in plasma microinstability analysis, Journal of Computational Physics 410 (2020) 109394.

[45] I.-G. Farcaş, A. D. Siena, F. Jenko, Turbulence suppression by energetic particles: a sensitivity-driven dimension-adaptive sparse grid framework for discharge optimization, Nuclear Fusion 61 (5) (2021) 056004. doi:10.1088/1741-4326/abecc8.
URL https://doi.org/10.1088/1741-4326/abecc8

[46] J. Lakhlili, O. Hoenen, O. O. Luk, D. P. Coster, Uncertainty quantification for multiscale fusion plasma simulations with vecma toolkit, in: V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos, J. Teixeira (Eds.), Computational Science – ICCS 2020, Springer International Publishing, Cham, 2020, pp. 719–730.

[47] P. Vaezi, C. Holland, An improved approach to uncertainty quantification for plasma turbulence validation studies, Fusion Sci. Technol. 74 (1-2) (2018) 77–88. doi:10.1080/15361055.2017.1372987.

[48] A. M. Dimits, G. Bateman, M. A. Beer, B. I. Cohen, W. Dorland, G. W. Hammett, C. Kim, J. E. Kinsey, M. Kotschenreuther, A. H. Kritz, L. L. Lao, J. Mandrekas, W. M. Nevins, S. E. Parker, A. J. Redd, D. E. Shumaker, R. Sydora, J. Weiland, Comparisons and physics basis of tokamak transport models and turbulence simulations, Physics of Plasmas 7 (3) (2000) 969–983.

[49] A. Owen, Y. Zhou, Safe and effective importance sampling, Journal of the American Statistical Association 95 (449) (2000) 135–143.
URL http://www.jstor.org/stable/2669533

[50] F. Zonta, L. Sanchis, E. Hirvijoki, the ASDEX Upgrade Team, A backward monte carlo method for fast-ion-loss simulations, Nuclear Fusion 62 (2) (2021) 026010. doi:10.1088/1741-4326/ac3a1b.
URL https://doi.org/10.1088/1741-4326/ac3a1b

[51] B. Peherstorfer, T. Cui, Y. Marzouk, K. Willcox, Multifidelity importance sampling, Computer Methods in Applied Mechanics and Engineering 300 (2016) 490–509. doi:https://doi.org/10.1016/j.cma.2015.12.002.
URL https://www.sciencedirect.com/science/article/pii/S004578251500393X

[52] B. Peherstorfer, B. Kramer, K. Willcox, Multifidelity preconditioning of the cross-entropy method for rare event simulation and failure probability estimation, SIAM/ASA Journal on Uncertainty Quantification 6 (2) (2018) 737–761. arXiv:https://doi.org/10.1137/17M1122992, doi:10.1137/17M1122992.
URL https://doi.org/10.1137/17M1122992

[53] A. Chaudhuri, B. Kramer, K. E. Willcox, Information reuse for importance sampling in reliability-based design optimization, Reliability Engineering & System Safety 201 (2020) 106853. doi:https://doi.org/10.1016/j.ress.2020.106853.
URL https://www.sciencedirect.com/science/article/pii/S0951832019301620

[54] L. W. Cook, J. P. Jarrett, K. E. Willcox, Generalized information reuse for optimization under uncertainty with non-sample average estimators, International Journal for Numerical Methods in Engineering 115 (12) (2018) 1457–1476. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nme.5904, doi:https://doi.org/10.1002/nme.5904.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.5904

[55] A. Giuliani, F. Wechsung, A. Cerfon, G. Stadler, M. Landreman, Single-stage gradient-based stellarator coil design: Optimization for near-axis quasi-symmetry, Journal of Computational Physics 459 (2022) 111147. doi:https://doi.org/10.1016/j.jcp.2022.111147.
URL https://www.sciencedirect.com/science/article/pii/S0021999122002091

[56] F. Wechsung, A. Giuliani, M. Landreman, A. Cerfon, G. Stadler, Single-stage gradient-based stellarator coil design: stochastic optimization, Nuclear Fusion 62 (7) (2022) 076034. doi:10.1088/1741-4326/ac45f3.
URL https://doi.org/10.1088/1741-4326/ac45f3

[57] C. D. Beidler, Y. I. Kolesnichenko, V. S. Marchenko, I. N. Sidorenko, H. Wobig, Stochastic diffusion of energetic ions in optimized stellarators, Physics of Plasmas 8 (6) (2001) 2731–2738. arXiv:https://doi.org/10.1063/1.1365958, doi:10.1063/1.1365958.
URL https://doi.org/10.1063/1.1365958

[58] J. Faustin, W. Cooper, J. Graves, D. Pfefferlé, J. Geiger, Fast particle loss channels in wendelstein 7-x, Nuclear Fusion 56 (9) (2016) 092006. doi:10.1088/0029-5515/56/9/092006.
URL https://doi.org/10.1088/0029-5515/56/9/092006

[59] R. White, S. Ethier, Particle resonances in stellarators, Physics of Plasmas 28 (9) (2021) 092503. arXiv:https://doi.org/10.1063/5.0060679, doi:10.1063/5.0060679.
URL https://doi.org/10.1063/5.0060679

[60] S. A. Lazerson, A. LeViness, J. Lion, Simulating fusion alpha heating in a stellarator reactor, Plasma Physics and Controlled Fusion 63 (12) (2021) 125033. doi:10.1088/1361-6587/ac35ee.
URL https://doi.org/10.1088/1361-6587/ac35ee

[61] H. E. Mynick, Transport optimization in stellarators, Physics of Plasmas 13 (5) (2006) 058102. arXiv:https://doi.org/10.1063/1.2177643, doi:10.1063/1.2177643.
URL https://doi.org/10.1063/1.2177643

[62] T. K. Mau, T. B. Kaiser, A. A. Grossman, A. R. Raffray, X. R. Wang, J. F. Lyon, R. Maingi, L. P. Ku, M. C. Zarnstorff, A.-C. Team, Divertor configuration and heat load studies for the aries-cs fusion power plant, Fusion Science and Technology 54 (3) (2008) 771–786. arXiv:https://doi.org/10.13182/FST08-27, doi:10.13182/FST08-27.
URL https://doi.org/10.13182/FST08-27

[63] J. B. Taylor, Equilibrium and stability of plasma in arbitrary mirror fields, The Physics of Fluids 7 (6) (1964) 767–773. arXiv:https://aip.scitation.org/doi/pdf/10.1063/1.1711283, doi:10.1063/1.1711283.
URL https://aip.scitation.org/doi/abs/10.1063/1.1711283

[64] P. Helander, D. J. Sigmar, Collisional transport in magnetized plasmas, Vol. 4, Cambridge university press, 2005.

[65] S. A. Lazerson, D. Pfefferlé, M. Drevlak, H. Smith, J. Geiger, S. Äkäslompolo, P. Xanthopoulos, A. Dinklage, O. Ford, P. McNeely, N. Rust, S. Bozhenkov, D. Hartmann, K. Rahbarnia, T. Andreeva, J. Schilling, C. Brandt, U. Neuner, H. Thomsen, R. C. Wolf, T. W.-X. Team, Modeling and measurement of energetic particle slowing down in wendelstein 7-x, Nuclear Fusion 61 (9) (2021) 096005. doi:10.1088/1741-4326/ac0771.
URL https://doi.org/10.1088/1741-4326/ac0771

[66] L. Ku, P. Garabedian, New classes of quasi-axisymmetric stellarator configurations, Fusion science and technology 50 (2) (2006) 207–215.

[67] Y. Liu, L. Li, A. Loarte, S. Pinches, A. Polevoi, Drift orbit islands of energetic particles due to 3d fields in ITER, Nuclear Fusion 61 (10) (2021) 106029. doi:10.1088/1741-4326/ac19fb.
URL https://doi.org/10.1088/1741-4326/ac19fb

[68] J. P. Freidberg, Plasma physics and fusion energy, Cambridge university press, 2008.

[69] M. Ohlberger, S. Rave, Reduced basis methods: Success, limitations and future challenges, Publishing House of Slovak University of Technology in Bratislava, 2016, pp. 1–12.

[70] A. Reiman, L. Ku, D. Monticello, S. Hirshman, S. Hudson, C. Kessel, E. Lazarus, D. Mikkelsen, M. Zarnstorff, L. Berry, et al., Recent advances in the design of quasiaxisymmetric stellarator plasma configurations, Physics of Plasmas 8 (5) (2001) 2083–2094.

[71] M. Zarnstorff, L. Berry, A. Brooks, E. Fredrickson, G. Fu, S. Hirshman, S. Hudson, L. Ku, E. Lazarus, D. Mikkelsen, et al., Physics of the compact advanced stellarator ncsx, Plasma Physics and Controlled Fusion 43 (12A) (2001) A237.

[72] G. Neilson, C. Gruber, J. H. Harris, D. Rej, R. Simmons, R. Strykowsky, Lessons learned in risk management on ncsx, IEEE transactions on plasma science 38 (3) (2010) 320–327.

[73] R. Dewar, S. Hudson, Stellarator symmetry, Physica D: Nonlinear Phenomena 112 (1-2) (1998) 275–280.

[74] C. Zhu, S. R. Hudson, Y. Song, Y. Wan, New method to design stellarator coils without the winding surface, Nuclear Fusion 58 (1) (2017) 016008. doi:10.1088/1741-4326/aa8e0a.
URL https://doi.org/10.1088/1741-4326/aa8e0a

[75] F. Najmabadi, A. R. Raffray, S. I. Abdel-Khalik, L. Bromberg, L. Crosatti, L. El-Guebaly, P. R. Garabedian, A. A. Grossman, D. Henderson, A. Ibrahim, T. Ihli, T. B. Kaiser, B. Kiedrowski, L. P. Ku, J. F. Lyon, R. Maingi, S. Malang, C. Martin, T. K. Mau, B. Merrill, R. L. Moore, R. J. P. Jr., D. A. Petti, D. L. Sadowski, M. Sawan, J. H. Schultz, R. Slaybaugh, K. T. Slattery, G. Sviatoslavsky, A. Turnbull, L. M. Waganer, X. R. Wang, J. B. Weathers, P. Wilson, J. C. W. III, M. Yoda, M. Zarnstorffh, The aries-cs compact stellarator fusion power plant, Fusion Science and Technology 54 (3) (2008) 655–672. arXiv:https://doi.org/10.13182/FST54-655, doi:10.13182/FST54-655.
URL https://doi.org/10.13182/FST54-655

[76] EPS Conference on Plasma Physics, Improvement of energetic particle confinement through stellarator optimization.

[77] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, Journal of Machine Learning Research 12 (2011) 2825–2830.