Rare Event Estimation Using Tensor Trains and Large Deviation Theory

Alex de Beer[*] Karina Koval[†] Robert Scheichl[‡,†] Tiangang Cui[*]

[*]School of Mathematics and Statistics, University of Sydney

[†]Interdisciplinary Center for Scientific Computing, Heidelberg University

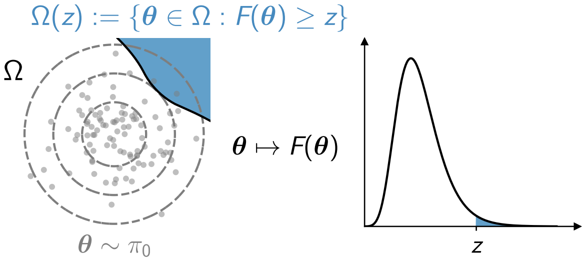
[‡]Institute for Mathematics, Heidelberg University

SIAM Conference on Uncertainty Quantification, March 2026

Overview

Rare Event Estimation

Rare events arise in a wide variety of contexts (*e.g.*, structural failure, volcanic eruptions, pandemics) and can have significant consequences.



We aim to estimate

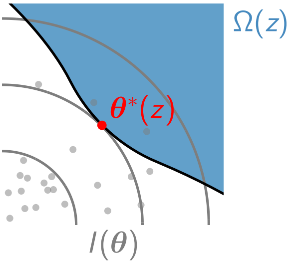
$$\mathbb{P}(z) := \int_{\Omega} \mathbb{I}_{\Omega(z)}(\boldsymbol{\theta}) \pi_0(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}.$$

Rare Event Estimation via Large Deviation Theory

Large deviation theory (LDT) concerns the asymptotic tail behaviour of probability distributions. Under certain conditions,* a variant of LDT states

$$\lim_{z\to\infty} \frac{-I(\boldsymbol{\theta}^*(z))}{\log \mathbb{P}(z)} = 1, \quad \text{where} \quad \boldsymbol{\theta}^*(z) := \arg\min_{\boldsymbol{\theta}\in\Omega(z)} I(\boldsymbol{\theta}),$$

and $I : \Omega \to \mathbb{R}$ denotes the *rate function* ($I(\boldsymbol{\theta}) = \frac{1}{2}||\boldsymbol{\theta}||^2$ if $\boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}_d, \boldsymbol{I}_d)$).



*Dematteis *et al.* (2019), Tong *et al.* (2021)

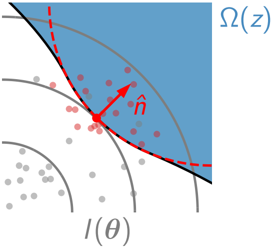
Large Deviation Theory-Based Importance Sampling

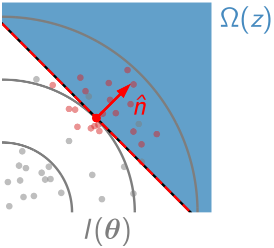
We can use LDT to design importance sampling algorithms that improve significantly on a naive Monte Carlo estimate.*

The eigendecomposition of the LDT Hessian, given by

$$\boldsymbol{H}_{\mathrm{LDT}}(\boldsymbol{\theta}) := (\boldsymbol{I} - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^\top)\nabla^2 F(\boldsymbol{\theta})(\boldsymbol{I} - \hat{\boldsymbol{n}}\hat{\boldsymbol{n}}^\top),$$

evaluated at $\boldsymbol{\theta}^*(z)$, provides insight into the directions in the parameter space that determine whether an event is rare.





*Tong and Stadler (2023)

Tensor-Based Approximations of Probability Densities

We can construct a tensor-based approximation of an arbitrary probability density function, π.* To do this, we

1. Form a tensor product discretisation of the parameter space. The corresponding **function evaluation tensor** $\mathbf{T} \in \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}$ has elements given by

$$\mathsf{T}[i_1, i_2, \ldots, i_d] = \pi(\theta_1^{(i_1)}, \theta_2^{(i_2)}, \ldots, \theta_d^{(i_d)}).$$

2. Select basis functions (splines, Chebyshev polynomials, etc.) to interpolate between the elements of $\mathbf{T}$.

We aim to construct an accurate approximation to $\mathbf{T}$ using as few evaluations of π as possible.

One way to do this is to use a tensor train decomposition. To form this we can use cross approximation,† which requires $\mathcal{O}(dnr^2)$ evaluations of π.

---

*Dolgov *et al.* (2020)

†Oseledets and Tyrtyshnikov (2010)

Extended Functional Tensor Train Approximations

To further reduce construction costs, we can consider an **extended** FTT approximation of π, where $\mathbf{T}$ is approximated in the format:*

$$\mathbf{T} \approx \mathbf{C} \times_1 \boldsymbol{F}_1 \times_2 \boldsymbol{F}_2 \times_3 \cdots \times_d \boldsymbol{F}_d,$$

where $\mathbf{C} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_d}$, $\boldsymbol{F}_k \in \mathbb{R}^{n_k \times R_k}$.

*Strössner *et al.* (2024)

Forming the Extended Functional Tensor Train

We start by forming a reduced basis for each mode of $\mathbf{T}$.

For $k \in \{1, 2, \ldots, d\}$:

1. Sample a sufficiently large random subset of mode-k fibres
2. Compute a reduced basis $\boldsymbol{U}_k \in \mathbb{R}^{n_k \times R_k}$ by taking a truncated SVD
3. Form an oblique projection $\boldsymbol{U}_k(\boldsymbol{P}_{I_k}^\top \boldsymbol{U}_k)^{-1}\boldsymbol{P}_{I_k}^\top$ (empirical interpolation),* where $\boldsymbol{P}_{I_k} \in \mathbb{R}^{n_k \times R_k}$ is a subset of columns of the identity matrix

The resulting approximation takes the form

$$\mathbf{T} \approx \underbrace{(\mathbf{T} \times_1 \boldsymbol{P}_{I_1}^\top \times_2 \cdots \times_d \boldsymbol{P}_{I_d}^\top)}_{\mathbf{C}=\mathbf{T}[I_1, I_2, \ldots, I_d]} \times_1 \underbrace{\boldsymbol{U}_1(\boldsymbol{P}_{I_1}^\top \boldsymbol{U}_1)^{-1}}_{\boldsymbol{F}_1} \times_2 \cdots \times_d \underbrace{\boldsymbol{U}_d(\boldsymbol{P}_{I_d}^\top \boldsymbol{U}_d)^{-1}}_{\boldsymbol{F}_d},$$

where $\mathbf{C}$ is a subtensor of $\mathbf{T}$ and can be approximated in TT format.

This can be significantly cheaper to construct than a direct TT factorisation if $R_k \ll n_k$.

*Chaturantabut and Sorensen (2010)

Sampling from Tensor-Based Density Approximations

We can derive the inverse Rosenblatt transport, $\mathcal{T} : \mathbb{R}^d \to \mathbb{R}^d$, which provides a coupling between $\hat{\pi}$ and an arbitrary product-form reference density ρ. This allows us to generate samples from $\hat{\pi}$.

ρ

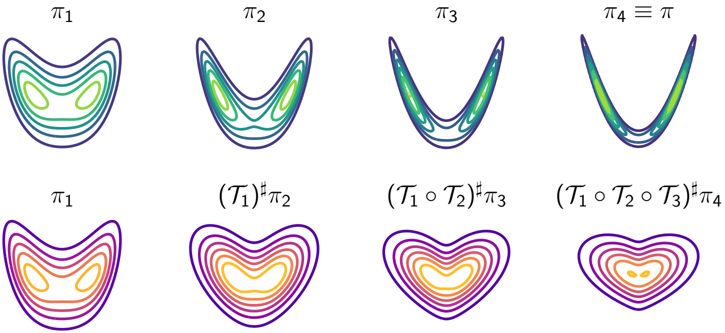
$$\boldsymbol{\theta} \mapsto \mathcal{T}(\boldsymbol{\theta})$$

$\hat{\pi}$

Deep Tensor-Based Density Approximations

If a density is highly concentrated or has a complex correlation structure, a layered composition of FTTs may approximate it more accurately.*

The construction of the FTTs is guided by a sequence of bridging densities, $\pi_1, \pi_2, \ldots, \pi_K \equiv \pi$.



$$\mathcal{T} = \mathcal{T}_1 \circ \mathcal{T}_2 \circ \mathcal{T}_3 \circ \mathcal{T}_4$$

*Cui and Dolgov (2022)

Rare Event Estimation Using Tensor Approximations

The optimal importance sampling density for estimation of a rare event probability takes the form

$$\pi^*(\boldsymbol{\theta}) := \frac{\mathbb{I}_{\Omega(z)}(\boldsymbol{\theta})\pi_0(\boldsymbol{\theta})}{\int_\Omega \mathbb{I}_{\Omega(z)}(\boldsymbol{\theta}')\pi_0(\boldsymbol{\theta}')\,\mathrm{d}\boldsymbol{\theta}'},$$

and gives a zero-variance estimate.

Cui *et al.* (2024) construct an FTT approximation, $\hat{\pi}^* \approx \pi^*$. Instead of approximating π^* directly, they approximate a smoothed version:

$$\pi^*_\gamma(\boldsymbol{\theta}) = \frac{f(\boldsymbol{\theta};\gamma)\pi_0(\boldsymbol{\theta})}{\int_\Omega f(\boldsymbol{\theta}';\gamma)\pi_0(\boldsymbol{\theta}')\,\mathrm{d}\boldsymbol{\theta}'}, \quad \text{where} \quad \lim_{\gamma\to\infty} f(\boldsymbol{\theta};\gamma) = \mathbb{I}_{\Omega(z)}(\boldsymbol{\theta}).$$

The smoothing parameter γ can also be used to define a sequence of bridging densities.

Our Contributions

Some areas we are interested in include

- Designing "preconditioning" strategies based on LDT, including working in terms of
  1. $\boldsymbol{u} = \boldsymbol{\theta} - \boldsymbol{\theta}^*(z)$, or
  2. $\boldsymbol{u} = \boldsymbol{\Phi}(z)^{-1}(\boldsymbol{\theta} - \boldsymbol{\theta}^*(z))$,

  where the columns of $\boldsymbol{\Phi}(z)$ are the eigenvectors of the LDT Hessian
- Investigating whether the optimal importance density can be approximated more accurately in extended FTT format
- Approximating the optimal importance density in the LDT subspace
- Approximating the optimal importance density in a subspace constructed adaptively using ideas from Bayesian inverse problems*
- Applying these ideas in settings where the distribution of the parameters is non-Gaussian (e.g., a Bayesian posterior)

*Uribe *et al.* (2021)

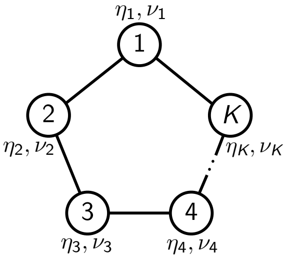
SIR Model

To illustrate these ideas, we consider a compartment susceptible-infectious-recovered (SIR) model governed by the dynamics

$$\frac{\mathrm{d}S_k}{\mathrm{d}t} = -\eta_k S_k I_k + \frac{1}{2}\sum_{j\in\mathcal{J}_k}(S_j - S_k),$$

$$\frac{\mathrm{d}I_k}{\mathrm{d}t} = \eta_k S_k I_k - \nu_k I_k + \frac{1}{2}\sum_{j\in\mathcal{J}_k}(I_j - I_k),$$

$$\frac{\mathrm{d}R_k}{\mathrm{d}t} = \nu_k I_k + \frac{1}{2}\sum_{j\in\mathcal{J}_k}(R_j - R_k),$$



$$\boldsymbol{\theta} := [\eta_1, \nu_1, \eta_2, \nu_2, \ldots, \eta_K, \nu_K]^\top$$

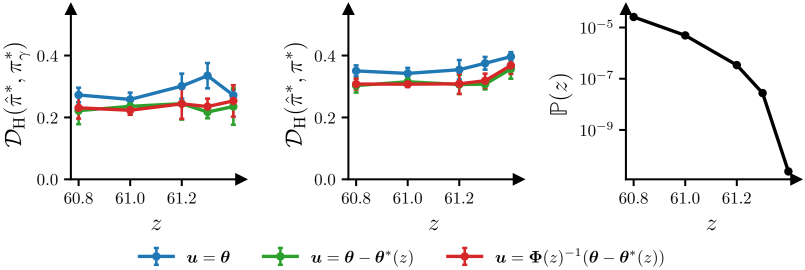
for $k \in \{1, 2, \ldots, K\}$, $t \in [0, T]$.

The parameter-to-event mapping is given by

$$F(\boldsymbol{\theta}) = \max_{t\in[0,T]} I_K(t; \boldsymbol{\theta}).$$

SIR Model: LDT-Based Preconditioning

We start by considering a setup with $K = 8$ compartments ($\boldsymbol{\theta} \in \mathbb{R}^{16}$), and investigate the impact of preconditioning strategies based on LDT.



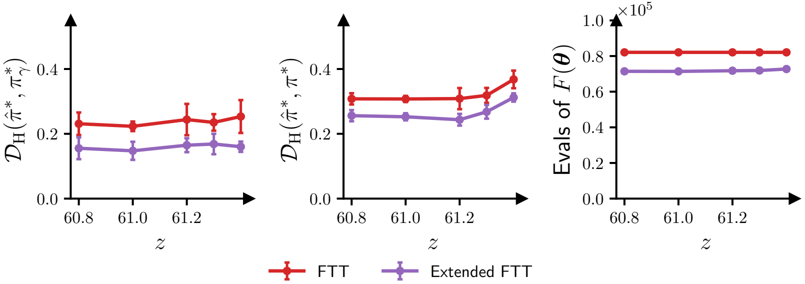
When approximating the optimal importance density, we use

- A tensor-product grid of size $n_k = 18$ in each dimension
- A fixed tensor train rank $r_k = 8$ in each dimension
- Piecewise linear basis functions
- Five bridging densities

SIR Model: Extended FTT

We now compare a standard FTT approximation of the optimal importance density to an extended FTT approximation.

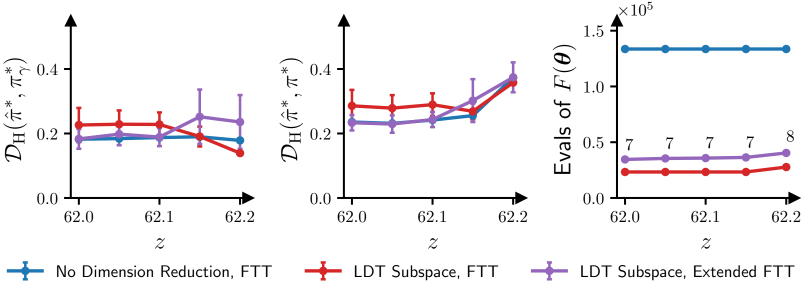
The extended FTT approximation uses a larger number of gridpoints than the FTT approximation ($n_k = 28$ vs $n_k = 18$), but the number of function evaluations required to construct each is comparable.



SIR Model: LDT-Based Dimension Reduction

We now consider a setup with $K = 16$ compartments ($\boldsymbol{\theta} \in \mathbb{R}^{32}$).

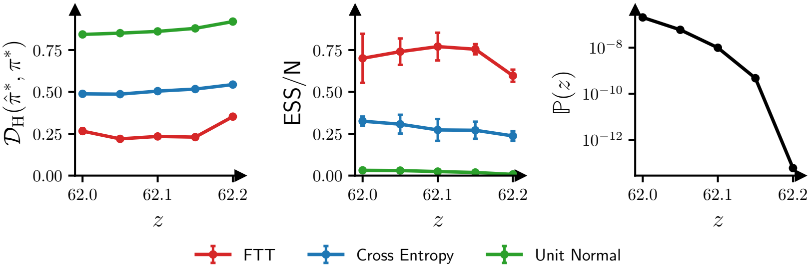
We compare the accuracy of an FTT approximation to the optimal importance density constructed without any dimension reduction, and some approximations constructed in the LDT subspace.



SIR Model: Comparison to Other Importance Densities

Finally, we compare the quality of the FTT-based approximation to the optimal importance density (constructed in the LDT subspace) to

1. The unit Gaussian distribution centred at the LDT point
2. A Gaussian proposal obtained using a cross entropy method in the LDT subspace



Contaminant Transport Model

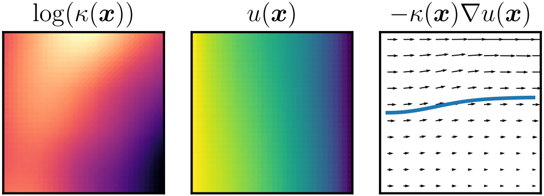
We now model a contaminant travelling through a fluid in a porous medium.

The fluid pressure, $u(\boldsymbol{x})$, depends on the unknown diffusivity, $\kappa(\boldsymbol{x})$:

$$-\nabla \cdot (\kappa(\boldsymbol{x})\nabla u(\boldsymbol{x})) = 0, \qquad \boldsymbol{x} \in [0,1]^2.$$

The trajectory of the contaminant is given by

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t} = -\kappa(\boldsymbol{x})\nabla u(\boldsymbol{x}), \qquad \boldsymbol{x}(0) = [0, 0.5]^\top.$$

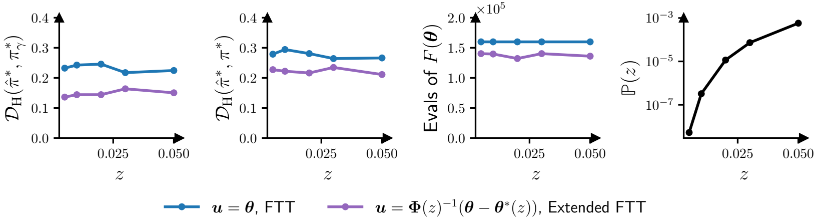


We aim to estimate the probability that the contaminant **breakthrough time** (to reach the right-hand boundary) is less than or equal to a specified threshold, z.

Contaminant Transport: LDT-Based Preconditioning

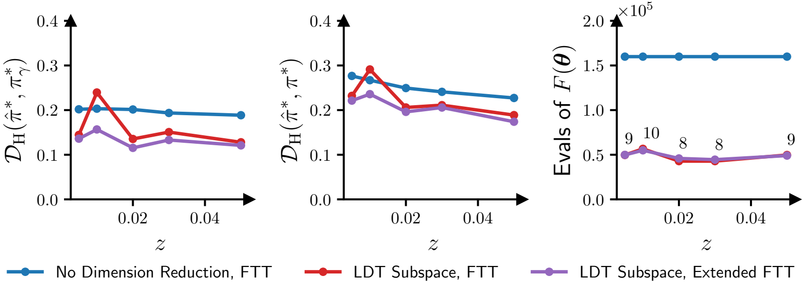
We parametrise the log-diffusivity $\log(\kappa(\boldsymbol{x}))$ using a KL expansion of a Gaussian random field with the Matérn covariance function ($\boldsymbol{\theta} \in \mathbb{R}^{25}$).

We start by comparing a standard FTT approximation of the optimal importance density to an extended FTT approximation that uses the LDT-based preconditioning techniques.



Contaminant Transport: LDT Subspace

We now compare the accuracy of an FTT approximation to the optimal biasing density constructed without any dimension reduction, and some approximations constructed in the LDT subspace.



Next Steps

Some conclusions:

- If the LDT subspace is available, constructing an FTT approximation in this subspace can be far more efficient than working in the full space
- Otherwise, an extended FTT approximation may be more efficient to construct than a standard FTT approximation

Ongoing work includes

- Reducing the complexity of approximating the optimal biasing density through preintegration*
- Designing alternative sequences of bridging densities that are easier to approximate using FTTs
- Applying these ideas to more challenging problems

*See, e.g., Griewank *et al.* (2018)

Software

All FTT approximations in this talk were constructed using the `deep_tensor` Python package.

`https://github.com/DeepTransport/deep-tensor-py`

References

Chaturantabut, S and Sorensen, DC (2010). Nonlinear model reduction via discrete empirical interpolation. *SIAM Journal on Scientific Computing* **32**, 2737–2764.

Cui, T and Dolgov, S (2022). Deep composition of tensor-trains using squared inverse Rosenblatt transports. *Foundations of Computational Mathematics* **22**, 1863–1922.

Cui, T *et al.* (2024). Deep importance sampling using tensor trains with application to a priori and a posteriori rare events. *SIAM Journal on Scientific Computing* **46**, C1–C29.

Dematteis, G *et al.* (2019). Extreme event quantification in dynamical systems with random components. *SIAM/ASA Journal on Uncertainty Quantification* **7**, 1029–1059.

Dolgov, S *et al.* (2020). Approximation and sampling of multivariate probability distributions in the tensor train decomposition. *Statistics and Computing* **30**, 603–625.

Griewank, A *et al.* (2018). High dimensional integration of kinks and jumps—smoothing by preintegration. *Journal of Computational and Applied Mathematics* **344**, 259–274.

Oseledets, I and Tyrtyshnikov, E (2010). TT-cross approximation for multidimensional arrays. *Linear Algebra and its Applications* **432**, 70–88.

Strössner, C *et al.* (2024). Approximation in the extended functional tensor train format. *Advances in Computational Mathematics* **50**, 54.

Tong, S and Stadler, G (2023). Large deviation theory-based adaptive importance sampling for rare events in high dimensions. *SIAM/ASA Journal on Uncertainty Quantification* **11**, 788–813.

Tong, S *et al.* (2021). Extreme event probability estimation using PDE-constrained optimization and large deviation theory, with application to tsunamis. *Communications in Applied Mathematics and Computational Science* **16**, 181–225.

Uribe, F *et al.* (2021). Cross-entropy-based importance sampling with failure-informed dimension reduction for rare event simulation. *SIAM/ASA Journal on Uncertainty Quantification* **9**, 818–847.