Thesis Proposal

Jordan R. Hall

Department of Mathematical and Statistical Sciences
University of Colorado Denver

Tuesday, December 4, 2018

Notation

- We define a model parameter space Λ with dimension N and a data space $\mathcal{D}$ with dimension M.
  - In most settings, Λ will be of higher dimension than $\mathcal{D}$.
- We define a parameter-to-data map $f : \Lambda \to \mathcal{D}$,
  - f may be polluted by noise.
  - ∇f may be inaccessible.
- We write $d = f(\lambda) \in \mathcal{D}$ to denote a particular datum corresponding to the evaluation of a point $\lambda \in \Lambda$.

The Stocastic Inverse Problem (SIP) [1][2][3]

Given:

- the *prior distribution*, $\pi_{\Lambda}^{\text{prior}}(\lambda)$: the prior or initial knowledge of the parameter space Λ
- the *observed distribution*, $\pi_{\mathcal{D}}(d)$: the uncertain state of knowledge of the observed data in $\mathcal{D}$.

The SIP is obtaining an *updated probability distribution*, $\pi_{\Lambda}^{\text{update}}(\lambda)$ combining the given prior information and the observed data.

[1] T. Butler et al, *Combining Pushforward Measures and Bayes' Rule to Construct Consistent Solutions to SIPs*

[2] Tarantola, Albert, *Inverse Problem Theory and Methods for Model Parameter Estimation*

[3] Stuart, Andrew., *Inverse problems: A Bayesian perspective*

The Forward Uncertainty Quantification (UQ) Problem

Given a probability distribution which is nonzero for every $\lambda \in \Lambda$, the *forward UQ problem* is finding the probability distribution of $f(\Lambda)$.

- The forward UQ problem is, in its own right, a nontrivial and important problem in UQ.
- The *classical Bayesian* or *statistical Bayesian* solution to the inverse problem is generally *not consistent*; i.e., it is not a *pull-back probability measure*.
  - This means the distribution of points drawn from the updated distribution $\pi_\Lambda^{\text{update}}$ and mapped through f, called the *push-forward* of $\pi_\Lambda^{\text{update}}$, is not equal to the observed probability distribution, $\pi_{\mathcal{D}}$.

- Data-Consistent Inversion (DCI) seeks an updated solution $\pi_\Lambda^{\text{update}}$ for which the push-forward exactly equals $\pi_{\mathcal{D}}$.
- To obtain such a solution, we must solve the forward UQ problem.
- We denote the solution to the forward UQ problem with $\pi_{\mathcal{D}}^{f(\Lambda)}(d)$.
- We present the data-consistent solution,

Data-Consistent Solution to the SIP [1]

$$\pi_\Lambda^{\text{update}}(\lambda) = \pi_\Lambda^{\text{prior}}(\lambda)\frac{\pi_{\mathcal{D}}(f(\lambda))}{\pi_{\mathcal{D}}^{f(\Lambda)}(f(\lambda))}. \tag{1}$$

[1] T. Butler et al, *Combining Pushforward Measures and Bayes' Rule to Construct Consistent Solutions to SIPs*

- Approximately solving the forward UQ problem needed to form (1) generally requires density estimation, which converges at best like Monte Carlo, $\sim \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$, where n represents the number of samples of f.
- In the case that f is a linear operator and the prior and observed densities are Gaussians, the solution to the SIP can be obtained exactly, and its mean is equivalent to the solution of a deterministic convex optimization problem.
  - The form of the objective function is different for statistical Bayesian inversion and DCI.

- Assume $\pi_\Lambda^{\text{prior}}$ and $\pi_\mathcal{D}$ are Gaussians with means λ_{prior} and d_{obs} and covariance matrices C_Λ and $C_\mathcal{D}$, respectively. If f is linear, then $f(\lambda) = A\lambda$. We define

The Classical Misfit Function [1]

$$S(\lambda) = \frac{1}{2}\left(\left|\left|C_\mathcal{D}^{-1/2}(A\lambda - d_{\text{obs}})\right|\right|_2^2 + \left|\left|C_\Lambda^{-1/2}(\lambda - \lambda_{\text{prior}})\right|\right|_2^2\right). \quad (2)$$

Comments

- The statistical mode of the posterior is called the maximum a posteriori point, or *MAP point*.
- The λ value that minimizes S is the MAP, λ_S^*, and the classical solution to the (linear) SIP is $\pi_\Lambda^{\text{post}}(\lambda) = c \cdot \exp(-S(\lambda))$.

[1] Tarantola, Albert, *Inverse Problem Theory and Methods for Model Parameter Estimation*

- S is reformulated so it is an objective function with a minimizer corresponding to the *data-consistent* MAP point.
- A *deregularization* term is appended so that if A is invertible, the regularization will be canceled. We define

The Data-Consistent Misfit Function [1]

$$T(\lambda) = \frac{1}{2}\left(\left|\left|C_{\mathcal{D}}^{-1/2}(A\lambda - d_{\text{obs}})\right|\right|_2^2 + \left|\left|C_{\Lambda}^{-1/2}(\lambda - \lambda_{\text{prior}})\right|\right|_2^2\right)$$
$$- \frac{1}{2}\left(\left|\left|C_{A}^{-1/2}(A(\lambda - \lambda_{\text{prior}}))\right|\right|_2^2\right), \qquad (3)$$

where $C_A = AC_\Lambda A^T$.

[1] Wildey et al, *A Consistent Bayesian Approach for SIPs Based on Push-forward Measures*

Comments

- If π_Λ and $\pi_\mathcal{D}$ are Gaussian, the data-consistent solution to our inverse problem is given exactly by $\pi_\Lambda^{\text{update}}(\lambda) = c \cdot \exp(-T(\lambda))$ or, in a more useful form, a Gaussian with mean λ_T^* and covariance matrix

$$C_\Lambda^U = (A^\top C_\mathcal{D}^{-1} A + C_\Lambda^{-1} - A^\top (C_A)^{-1} A)^{-1}$$

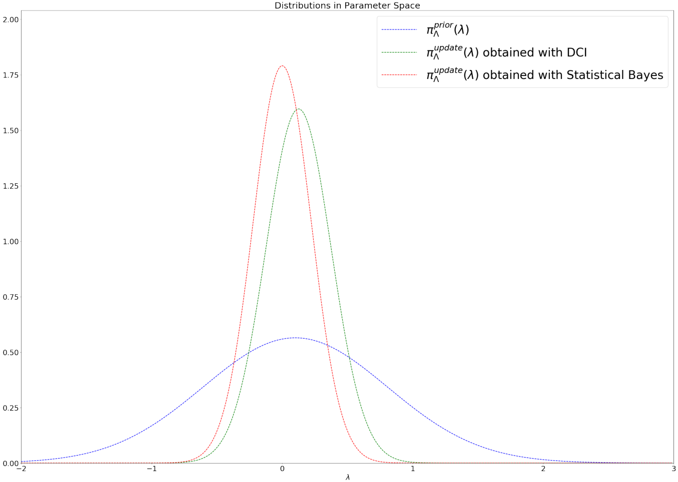
- The deregularization term in (3) ensures the updated distribution on Λ will only be regularized in directions uninformed by data.

Example 1.

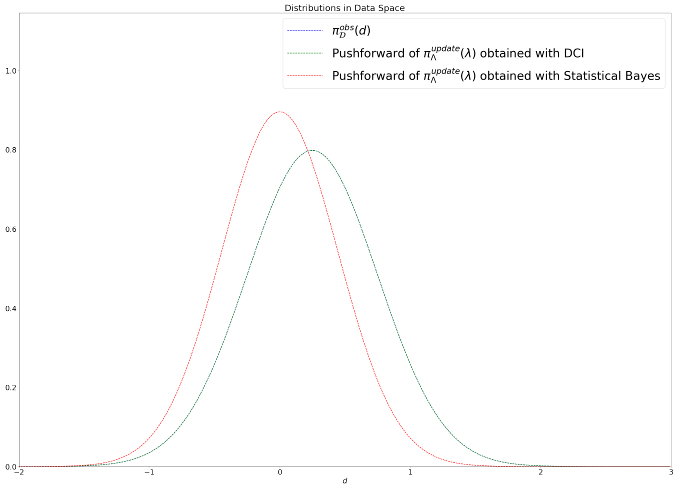
Let $f(\lambda) = 2\lambda$, $\lambda_{\text{prior}} = 0.1$, $C_\Lambda = [0.5]$, $d_{\text{obs}} = 0.25$, and $C_{\mathcal{D}} = [0.25]$; note, $N = M = 1$. We find

$$S(\lambda) = 2((2\lambda - 0.25)^2 + (\lambda - 0.1)^2) \quad \text{and} \quad T(\lambda) = 2(2\lambda - 0.25)^2.$$

- S is minimized by $\lambda_S^* = 3/25$.
- With Gaussian assumptions on the prior and data, we have $\pi_\Lambda^{\text{post}} \sim \exp(-S(\lambda))$ which is a $N(\lambda_S^*, 1/18)$.
- Notice $f(\lambda_S^*) = 6/25 \neq d_{\text{obs}}$.
- T is minimized by $\lambda_T^* = 1/8$.
- With Gaussian assumptions on the prior and data, we have $\pi_\Lambda^{\text{update}} \sim \exp(-T(\lambda))$ which is a $N(\lambda_T^*, 1/16)$.
- Notice $f(\lambda_T^*) = d_{\text{obs}}$.



Distributions in parameter space from Example 1.



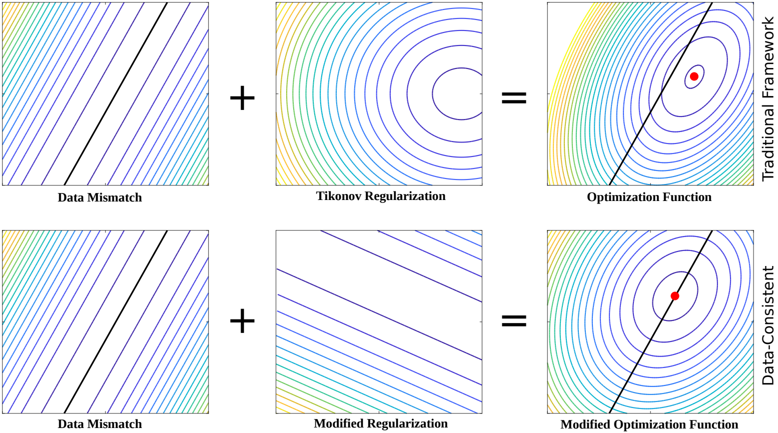
Distributions in data space from Example 1.

Example 2. [1]

Let $f(\lambda) = 2\lambda_1 - \lambda_2$, $\lambda_{\text{prior}} = (0.1 \quad 0.2)^\top$, $C_\Lambda = \text{diag}[0.5, 0.25]$, $d_{\text{obs}} = 0.1$, and $C_\mathcal{D} = [0.25]$. ($N = 2$ and $M = 1$.)

- For any $\lambda \in \Lambda$, $f(\lambda) = d = 2\lambda_1 - \lambda_2$. Since there is just a single d_{obs}, we have $0.1 = 2\lambda_1 - \lambda_2$.
- S is minimized by $\lambda_S^* = (7/50, 19/100)$. With Gaussian assumptions, we have $\pi_\Lambda^{\text{post}} \sim \exp(-S(\lambda))$ with mean λ_S^*. Notice $f(\lambda_S^*) = 9/100 \neq d_{\text{obs}}$.
- In the data-consistent formulation, we have T minimized by $\lambda_T^* = (13/90, 17/90)$. Notice $f(\lambda_T^*) = 9/90 = d_{\text{obs}}$. With Gaussian assumptions, we have $\pi_\Lambda^{\text{update}} \sim \exp(-T(\lambda))$ with mean λ_T^*.

[1] Wildey et al, *A Consistent Bayesian Approach for SIPs Based on Push-forward Measures*



This figure [1] shows the process of obtaining the statistical/classical Bayesian solution and data-consistent solution in Example 2.

[1] Wildey et al, *A Consistent Bayesian Approach for SIPs Based on Push-forward Measures*

- We consider functions $f : \Lambda \to \mathcal{D}$ where $\dim(\Lambda) = N$ is large and $\dim(\mathcal{D}) = M$ is such that $M < N$ or $M << N$.
  - Functions of interest may represent postprocessed quantities from the solution of complex physical models.
- It is not often that every parameter has equal impact on function values – usually some parameters matter more than others.
- The dimension reduction techniques considered seek to explain outputs $f(\Lambda)$ in an *active subspace* $\mathcal{A} \subset \Lambda$ for which $\dim(\mathcal{A}) < N$.
  - Many common uses (e.g., statistics) of f involve integration over Λ and are subject to the *curse of dimensionality*.
  - This forces the use of Monte Carlo or Quasi-Monte Carlo methods.
  - A lower-dimensional representation of f may enable faster methods.

- $\nabla f(\lambda) \in \Lambda$ is a column vector containing the N partial derivatives of f, which for this discussion we assume exist, and are square integrable in Λ equipped with some probability density that is positive everywhere in Λ and 0 otherwise.
  - We consider $\pi_\Lambda^{\text{prior}}(\lambda)$, the density describing our prior state of knowledge, which we abbreviate as π_Λ.
- For convenience, one transforms inputs λ to the origin with some fixed variance, typically so that $\lambda \in [-1, 1]^N$. We define

Covariance in Gradient Space [1]

$$W = \int_\Lambda \nabla f(\lambda) \nabla f(\lambda)^\top \pi_\Lambda(\lambda) d\lambda, \tag{4}$$

which is an $N \times N$ symmetric positive semi-definite matrix.

[1] Constantine, *Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies*

- Interpreting W as a certain covariance structure over Λ leads one to the idea of computing the Singular Value Decomposition of W,

Singular Value Decomposition (SVD) of W

$$W = U\Sigma V^*, \tag{5}$$

where U is $N \times N$ unitary, Σ is $N \times N$ diagonal with the singular values of W along its diagonal, and V^* is $N \times N$ unitary.

- We plot the singular values, $\{\sigma_i\}_{i=1}^n$ and seek a drop-off in magnitude between some pair of singular values, σ_j and σ_{j+1}. The active subspace is the span of $u_1, \ldots, u_j$, which are the first j columns of U, the left singular vectors of W.

- For a point $\lambda \in \Lambda$, we define

Projection into $\mathcal{A}$, *active variables*

$$\mathcal{P}_{\mathcal{A}}(\lambda) = \sum_{i=1}^{j} \left(u_i^T \lambda\right) u_i, \tag{6}$$

which is the projection of λ in the active directions of f.

- We have arrived at the property that

Resolution of f in $\mathcal{A}$

$$f\left(\mathcal{P}_{\mathcal{A}}(\lambda)\right) \approx f(\lambda). \tag{7}$$

- Finding an active subspace requires forming an approximation to W via Monte Carlo. Here we consider techniques in [1] [2].
- We let $D_S = \{(\lambda_i, f(\lambda_i))\}_{i=1}^{S}$, which is a set of S pairs of samples $\lambda_i \in \Lambda$ and their function values.
- One may use D_S to approximate ∇f. We denote each

  estimation to $\nabla f(\lambda_i) \approx \widehat{\nabla f}(\lambda_i)$.

- We form the $N \times S$ matrix $\tilde{W}$ (which we present as $\tilde{W}^\top$)

Monte Carlo Approximation to W [1]

$$\tilde{W}^\top := \left[\widehat{\nabla f}(\lambda_1) \cdots \widehat{\nabla f}(\lambda_S) \right]. \tag{8}$$

[1] Russi, *UQ with Experimental Data and Complex System Models*

[2] Constantine et al, *Computing Active Subspaces Efficiently with Gradient Sketching*

- Forming the SVD of $\tilde{W}$, $\tilde{W} = \tilde{U}\tilde{\Sigma}\tilde{V}^*$, we search for a drop off in the magnitude of the singular values $\{\tilde{\sigma}_i\}_{i=1}^{S}$. Assuming such a drop off occurs for an index $j : 1 < j < S$, we have the j corresponding left singular vectors,$\tilde{u}_1, \ldots, \tilde{u}_j$.
- Then we define

Monte Carlo approximation to $\mathcal{A}$

$$\mathcal{A}\,(f; D_S) := \text{span}\{\tilde{u}_1, \ldots, \tilde{u}_j\},$$

the active subspace of f with respect to the samples D_S.

- For low dimensional $\mathcal{A}$, we may check $f(\mathcal{P}_{\mathcal{A}}(\lambda)) \approx f(\lambda)$ in a *sufficient summary plot*, where we plot active variables against function values.

- Many important physical systems possess turbulent or chaotic behavior.
- The physical state of the system $u(x, \lambda)$ and the corresponding parameter to observable map $f(u(x, \lambda))$ may be modeled as a stochastic process, or as a deterministic function with additive or multiplicative noise.
    - In this setting, the efficient extraction of accurate gradients of f in parameter space is a challenging undertaking, as popular techniques based on linearization, including adjoint methods, are inaccurate [1] [2].
    - The finite-difference approximation of ∇f_{Λ} involve $N = \dim\Lambda$ additional, usually nonlinear model solves for the physical system state $u(x, \lambda_i + \delta\lambda_i)$, and may be greatly polluted by the noise in f.

---

[1] Lea, *Sensitivity analysis of the climate of a chaotic system*

[2] Qiqi, *Least Squares Shadowing sensitivity analysis of chaotic limit cycle oscillations*

- We are interested in derivative-free optimization (DFO) algorithms suited for additive and multiplicative noise. These techniques only require evaluations of the noisy model and random draws from a normal distribution.
  - In particular, we consider the STep-size Approximation in Randomized Search (STARS) algorithm [1].
- The smoothing factor and step size in STARS depend on scale factors of the L_1 Lipschitz constant of f. It is of interest to obtain estimates of L_1, which is not straightforward in a gradient-free setting. We refer to [2] [3] for Lipschitz constant learning.
- We note that **the convergence of STARS is dimension dependent.**

---

[1] Chen and Wild, *Randomized DFO of Noisy Convex Functions*

[2] Jan-Peter Calliess, *Lipschitz optimisation for Lipscitz interpolation*

[3] Kvasov and Sergeyev, *Lipschitz gradients for global optimization in a one-point-based partitioning scheme*

» We consider the problem

Optimization Under Additive Uncertainty [1]

$$\min_{\lambda \in \mathbb{R}^N} \quad \mathbb{E}\left[f(\lambda) + \nu(\lambda;\epsilon)\right], \tag{9}$$

» where:

(i.) $f : \mathbb{R}^N \to \mathbb{R}$ is convex;
(ii.) ϵ is a random variable with probability density $P(\epsilon)$;
(iii.) for all λ the additive noise model ν is independent and identically distributed, has bounded variance σ_a^2, and is unbiased; i.e., $\mathbb{E}_\epsilon(\nu(\lambda;\epsilon)) = 0$.

[1] Chen and Wild, *Randomized DFO of Noisy Convex Functions*

Stepsize h

$$h = (4L_1(N+4))^{-1}$$

where N is the dimension and L_1 is the global Lipschitz constant for $\|\nabla f\|$.

Smoothing factor μ^*

$$\mu^* = \left[\frac{8\sigma_a^2 N}{L_1^2(N+6)^3}\right]^{\frac{1}{4}}$$

where σ_a^2 is the variance of the additive noise.

Algorithm 1: *Minimization of $f+\epsilon$ via STARS*[1].

1: Define: $\left(\texttt{maxit}; \lambda^{(0)}; f_0 := f(\lambda^{(0)}) + \epsilon_0; \mu^*; h\right)$. Set `i=1`.

2: Draw a random $N \times 1$ vector r^i, where $r^i_j \sim N(0,1)$ for $j = 1, \dots, N$ and draw ϵ_i.

3: Evaluate $g_i := f(\lambda_{i-1} + \mu^* u_i) + \epsilon_i$.

4: Set $d_i := \dfrac{g_i - f_{i-1}}{\mu^*} u_i$.

5: Set $\lambda_i = \lambda_{i-1} - h \cdot d_i$.

6: Evaluate $f_i := f(\lambda_i) + \epsilon_i$; set `i=i+1`; return to 2.

7: Terminate when `i=maxit`.

[1] Chen and Wild, *Randomized DFO of Noisy Convex Functions*

- Recall for a linear f where $f(\lambda) = A\lambda$, the deregularization term in (3) is

Deregularization Term

$$\left|\left|C_A^{-1/2}(A(\lambda - \lambda_{\text{prior}})\right|\right|_2^2, \tag{10}$$

where $C_A = AC_\Lambda A^\top$.

Research Question

For a nonlinear f, inversion via convex optimization may be performed if f is linearized and Gaussian assumptions are met.

- What form should the deregularization term take to roughly preserve data consistency?
- Can we construct error bounds based on local linearization error?

For a linear map f and Gaussian knowledge state, the optimization solution of a SIP will be exact.

Research Questions

- To what extent does additive or multiplicative noise alter DCI via optimization?
- We investigate how using DFO to solve for a data-consistent MAP point compares to standard DCI methods (density estimation) in this setting.

We are interested in several problems that involve sampling Λ and evaluating f including solving the forward problem, finding an active subspace via Monte Carlo, and performing DFO.

Research Questions

- We are interested in comparing active subspaces obtained from sampling f with few samples, or with samples generated from some other process, such as a DFO algorithm.
- We are interested in learning Lipschitz constants [1] [2] from sampling, where samples are few and potentially generated by DFO.

[1] Jan-Peter Calliess, *Lipschitz optimisation for Lipscitz interpolation*

[2] Kvasov and Sergeyev, *Lipschitz gradients for global optimization in a one-point-based partitioning scheme*

» We are interested in investigating the effectiveness of only optimizing f in its active variables.

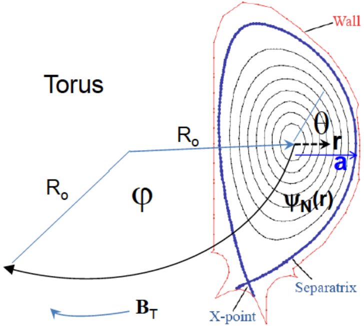
» The most compelling approach we have observed is to modify the DFO algorithm discussed above to only take random walks in directions lying in $\mathcal{A}$.

Modified DFO Algorithm

- Obtain the first j singular unit vectors $u_1, \ldots, u_j$ corresponding to the SVD of $\hat{W}$.
- Take j draws from a specified normal distribution, which we denote with $s_i \sim N(\mu, \sigma^2)$.
- Form the random vector v for the k-th step in a DFO algorithm as $v^{(k)} = \sum_{i=1}^{j} s_i u_i$.

» Other weighting schemes on the u_i's could be considered, such as weights given by singular values.

Tokamak Geometry

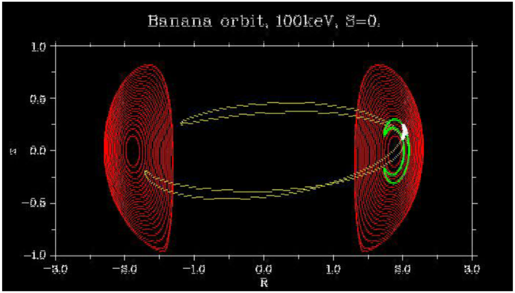


Poloidal cross section at a constant toroidal angle

» Normalized flux coordinate ψ_N: 0 at r, 1 at $\frac{r}{a} = 1$.

Transport in Tokamaks

Transport in tokamak plasmas can be strongly driven by a combination of *neoclassical* effects and plasma *microturbulence*.



Neoclassical theory predicts complex particle motion

» Neoclassical transport: simplified Vlasov-Boltzman equation coupled with Maxwell's equations [1].

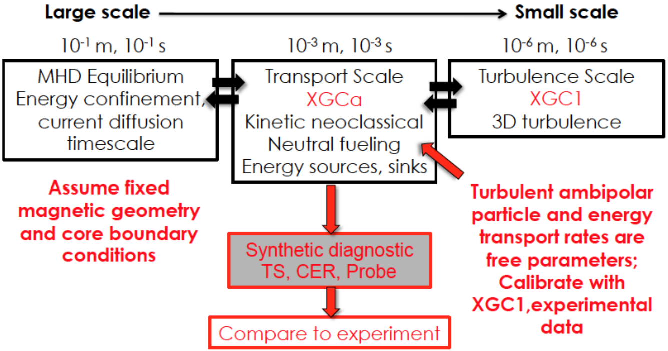
[1] Wesson, *Tokamaks*

Simulating the microturbulence in the plasma involves very expensive kinetic simulations, especially when simulating the plasma edge [1].

- An alternate approach to simulate microturbulence is by fitting an anomalous diffusion model.
  - Fit using experimental data
  - Direct Numerical Simulation (DNS) approach: Fit using high-fidelity simulation data.
  - Deterministic efforts in this setting include [2].
- This approach is similar to common modeling techniques in parameterizing micro-scale effects in macro/meso-scale systems (e.g., turbulence models, sub-grid models).

[1] S. Ku et al, *A new hybrid-Lagrangian numerical scheme for gyrokinetic simulation of tokamak edge plasma*

[2] Battaglia et al, *Kinetic neoclassical transport in the H-mode pedestal*



Research Question

We propose to investigate the data-consistent calibration of an anomalous diffusion model in an axisymmetric setting (XGCa) to high-fidelity experimental or simulation data (XGC1) [1].

[1] S. Ku et al, *A new hybrid-Lagrangian numerical scheme for gyrokinetic simulation of tokamak edge plasma*

» The dimensionless, axisymmetric Grad-Shafronov equation:

The Grad-Shafronov Equation [1]

$$r\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial\psi_n}{\partial r}\right)+\frac{\partial^2\psi_n}{\partial z^2}=-\frac{\alpha^2}{2}\frac{df^2}{d\psi_n}-r^2\alpha^2\frac{dp}{d\psi_n}. \tag{11}$$

» In equilibrium fitting to data [2]: p and f^2 are expanded as basis functions in ψ_n, with the constraint that p and f^2 vanish at the boundary.

Research Question

The updated PDF π_Λ^u is important to prediction of plasma profiles. **DCI problem**:Find the updated PDF of the uncertain parameters in the expansions of p and f^2.

[1] Takeda, T. and Tokuda, S., *Computation of MHD equilibrium of tokamak plasma*

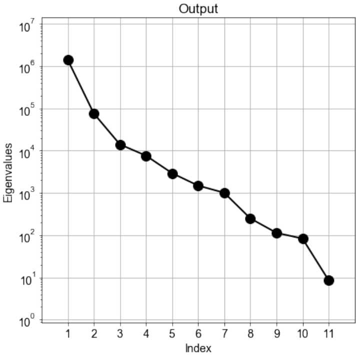
[2] Lao et al, *Reconstruction of current profile parameters and plasma shape in tokamaks*

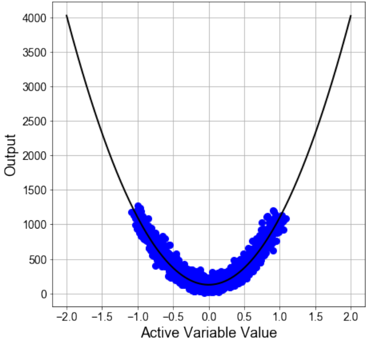
Example 3.

Let $\Lambda = [-1, 1]^{11}$ and define

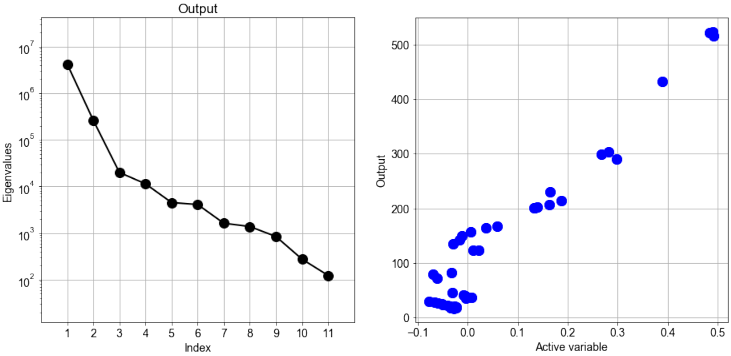
$$f(\lambda) = \sum_{i=0}^{10} 2^{(-1)^i i} \lambda_i^2 + \epsilon(\lambda),$$

where $\epsilon(\lambda)$ is a draw of additive noise corresponding to the input λ; here, we take draws of ϵ of order 10^{-4}. We see that $\mathcal{D} = [0, 2^{10}]$ and $N = 11$, $M = 1$. Note that the minimum of f is given by $0 \in \Lambda$. Here, as i increases, terms in f become either more important or less important, depending on whether i is even or odd.

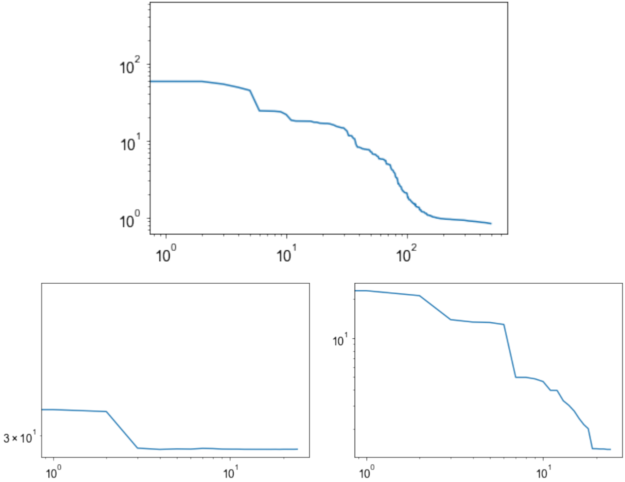




Left: A plot of the eigenvalues of the matrix $\hat{W}$ formed from 1000 Monte Carlo samples in Λ. We see one dominant eigenvalue on the order of 10^6. Right: A sufficient summary plot where all 1000 samples are projected into $\mathcal{A}$ and plotted against their function values; a quadratic surrogate fits the projected data with $R^2 \approx 0.9$.



Left: A plot of the eigenvalues of the matrix $\hat{W}$ formed from using 100 DFO iterates as samples in Λ. We again see one dominant eigenvalue between orders 10^6 and 10^7. Right: A sufficient summary plot where all 100 samples are projected into $\mathcal{A}$ and plotted against their function values.



Top: 500 iterations of standard DFO; Bottom Left: 25 iterations of minimizing f with DFO along the λ_8 axis; Bottom Right: 25 iterations of minimizing f with DFO along the λ_6 axis.

We must consider ways to express the data-consistent deregularization for nonlinear f. This is a theoretical task which will require a closer literature review and heavy collaboration with advisers.

Research Plan

- Reformulate problem as minimization of the statistical Bayesian misfit constrained by data-consistency, which will impose a nonlinear equality constraint.

Under assumptions of Gaussian priors and data and a linear f, the solution of our SIP is exactly determined by optimization methods. How do multiplicative or additive noise models on f perturb the solution of the SIP?

Research Plan

- We propose to analyze additive and multiplicative noise in the context of optimization methods for DCI.
- We consider the impact of this noise on MAP points and updated priors, especially when optimization is performed in a derivative-free setting.

We propose to investigate the approximate solution of data-consistent inverse problems in tokamak plasma modeling via optimization approaches.

Derivative Free Optimization

- Initial model functions + noise
- DCI of a subgrid large eddy simulation(LES) model for isotropic turbulent flow.
- DCI of an anomalous diffusion model for gyrokinetic turbulence in XGCa.

Gradient-based methods

- Approximate DCI for a model elliptic problem with a parameterized forcing function, with parameter gradients obtained via adjoint methods.
- Data-consistent MHD equilibria (Grad-Shafronov)

- Currently, only some of the software used to produce results in this paper are publicly available on GitHub.com.
- Some of the algorithms used here and other algorithms that are of interest are within open-source packages available online [1] [2].
- Other schemes considered here make modifications to given algorithms and remain under development.
- A major goal of this thesis proposal will be producing well-documented, open-source software complete with python Jupyter Notebooks containing illustrative, replicable examples.

[1] T. Butler et al
[2] Constantine et al

We outline a rough timeline for the remaining 2 years in a 5.5 year plan.

- Clean existing algorithms and examples, generate richer research results related to DFO and active subspaces, and build a model inverse problem for investigation. (Dec 2018/Jan 2019)
- On RA for Spring 2019. (Jan 2019-May 2019)
- Write and present MS-level results. (Feb/Mar 2019)
- Work on theoretical formulation of deregularization for nonlinear f. Work on generating notebooks and examples. (Ongoing/Spring and Summer 2019)
- Summer research; begin writing thesis; summer school/internship/conferences. (Jun/Jul/Aug 2019)
- Fall 2019 - Writing phase, revisions.
- Spring 2020 - Final revisions, software, documentation.
- Summer 2020 Internship/collaborations, publishing, formatting.
- Defend thesis Summer 2020 or early Fall 2020.

- Battaglia, D. J. and Burrell, K. H. and Chang, C. S. and Ku, S. and deGrassie, J. S. and Grierson, B. A. “Kinetic neoclassical transport in the H-mode pedestal." Physics of Plasmas,Volume 21, No. 7. 2014.
- T. Butler and J. Jakeman and T. Wildey. “Combining Push-Forward Measures and Bayes’ Rule to Construct Consistent Solutions to Stochastic Inverse Problems." SIAM Journal on Scientific Computing, Volume 40, No. 2, pp. A984-A1011, 2018.
- Jan-Peter Calliess. “Lipschitz optimisation for Lipschitz interpolation." In 2017 American Control Conference (ACC 2017), Seattle, WA, USA, May 2017.
- Chen and Wild. “Randomized Derivative-Free Optimization of Noisy Convex Functions." Funded by the Department of Energy. 2015.
- Constantine, Paul G. “Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies." SIAM, 2015.
- Constantine, Eftekhari, Wakin. “Computing Active Subspaces Efficiently with Gradient Sketching." Conference paper, 2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP).
- Kvasov and Sergeyev. “Lipschitz gradients for global optimization in a one-point-based partitioning scheme." Journal of Computational and Applied Mathematics. Volume 236, Issue 16, pp. 4042-4054. 2012.
- S. Ku and R. Hager and C.S. Chang and J.M. Kwon and S.E. Parker. “A new hybrid-Lagrangian numerical scheme for gyrokinetic simulation of tokamak edge plasma." Journal of Computational Physics, Volume 315, pp. 467-475. 2016.
- Lao, L.L, St. John, H, R.D. Stambaugh, A.G. Kellman, and Pfeiffer, W.,“Reconstruction of current profile parameters and plasma shapes in tokamaks”, Nuclear Fusion, Volume 25,No. 11,pp. 1611,1985.

- Lea, Daniel J. and Allen, Myles R. and Haine, Thomas W. N. “Sensitivity analysis of the climate of a chaotic system." Tellus A, Volume 52, No. 5, pp. 523-532. 2000.
- Russi, Trent M. “Uncertainty Quantification with Experimental Data and Complex System Models." Dissertation, University of California Berkeley. 2010.
- Smith, Ralph.“Uncertainty Quantification: Theory, Implementation, and Applications.” SIAM, 2013.
- Stuart, Andrew. “Inverse problems: A Bayesian perspective." Acta Numerica, volume 19, pp. 451-559. 2010.
- Tarantola, Albert. “Inverse Problem Theory and Methods for Model Parameter Estimation." SIAM. 2005.
- Takeda, T. and Tokuda S.,“Computation of MHD equilibrium of tokamak plasma”, Journal of Computational Physics,93,1,1 - 107,1991.
- Qiqi Wang and Rui Hu and Patrick Blonigan. “Least Squares Shadowing sensitivity analysis of chaotic limit cycle oscillations." Journal of Computational Physics, Volume 267, pp. 210-224. 2014.
- Wesson, J. “Tokamaks." “Oxford University Press, 4th edition." 2011.
- Wildey, T., Butler, T., Jakeman, J., Walsh, S. " A Consistent Bayesian Approach for Stochastic Inverse Problems Based on Push-forward Measures." SAND2017-3436PE. 2017.