Optimal Design Under Uncertainty

Edward R. Dougherty

Department of Electrical and Computer Engineering
Texas A&M University

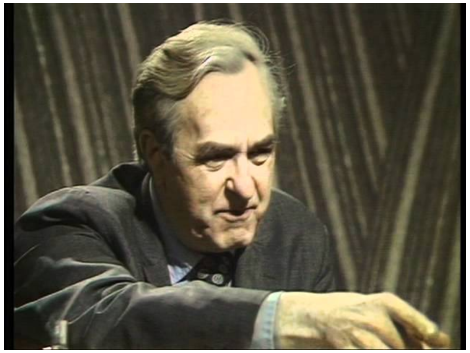
Bacon on Planned Experiments

- **Francis Bacon** (*Novum Organum,* 1620): "There remains simple experience which, if taken as it comes, is called accident; if sought for, experiment. But this kind of experience is…a mere groping, as of men in the dark… But the true method of experience, on the contrary, first lights the candle, and then by means of the candle shows the way; commencing as it does with experience duly ordered and digested, not bungling or erratic, and from it educing axioms, and from established axioms again new experiments."

Texas A&M GSPLAB

Models Precede Data

- **William Barrett**: "The scientist's mind is not a passive mirror that reflects the facts as they are in themselves (whatever that might mean); the scientist constructs models, which are not found among the things given him in his experience, and proceeds to impose those models upon Nature. And he must often construct those models conceptually before they are translated at any point into the material constructions of his apparatus in the laboratory."



An Experiment is a Question

- **Hans Reichenbach** (*Rise of Scientific Philosophy*): “An experiment is a question addressed to Nature….As long as we depend on the observation of occurrences not involving our assistance, the observable happenings are usually the product of so many factors that we cannot determine the contribution of each individual factor to the total result.”

Texas A&M GSPLab

Foolish Questions Yield Foolish Answers

- **Norbert Wiener**: "An experiment is a question. A precise answer is seldom obtained if the question is not precise; indeed, foolish answers – i.e., inconsistent, discrepant or irrelevant experimental results – are usually indicative of a foolish question."

Models Depend on Questions Asked

- **Werner Heisenberg**: “The most important new result of nuclear physics was the recognition of the possibility of applying quite different types of natural laws, without contradiction, to one and the same physical event. This is due to the fact that within a system of laws which are based on certain fundamental ideas only certain quite definite ways of asking questions make sense, and thus, that such a system is separated from others which allow different questions to be put.

Answers Without Questions

- **Hannah Arendt**: "The experiment "being a question put before nature" (Galileo), the answers of science will always remain replies to questions asked by men; the confusion in the issue of objectivity was to assume that there could be answers without questions and results independent of a question-asking being."

The Basic Engineering Problem

- Given a physical system, an objective, and a cost function quantifying the achievement of the objective, find an operator from a class of operators to minimize the cost.
  - Control theory: find an optimal controller to alter the behavior of the system – for instance, a drug to alter gene regulation.
  - Image processing: find an optimal filter to alter the properties of the image – for instance, reduce noise to improve recognition.
  - Classification: find an optimal classifier to make decisions – for instance, decide the stage of a cancer.

Texas A&M GSPLab

Our Problem

- Our problem is that we do not have an accurate model for the underlying system.
  - Insufficient scientific knowledge and insufficient data.
- In this case there are two possibilities:
  - **Optimal**: Use whatever knowledge we have plus data and find an optimal operator relative to the prior knowledge and the data.
  - **Ad hoc**: Assume a form for the operator and use data to construct an operator that is hopefully close to optimal – study its properties.
- For classification, operator forms are classification rules.
  - LDA, SVM, Neural network, etc.

Classification

- For classification the scientific model is the feature-label distribution, the objective is to choose a class, the cost is the classification error, and a **Bayes classifier** minimizes the error, the minimum being the **Bayes error**.
  - There can be many Bayes classifiers.
  - The Bayes error is intrinsic to the feature-label distribution.
- For classification, operator forms are classification rules.
  - LDA, SVM, Neural network, etc.
  - These rules are ad hoc in the sense that they are not derived via optimality, but rather constructed via heuristics.
  - Their goodness is evaluated separately from their definition.

Robust Operators

- If the scientific model is uncertain but its structure and some of its parameters are known, then the true model belongs to an *uncertainty class* Θ of models determined by a vector θ of unknown parameters.
- There is a cost function C and a class Ψ of operators.
- An *intrinsically Bayesian robust* (IBR) operator minimizes the expected value of the cost with respect to a prior probability distribution $\pi(\theta)$ over Θ.
  - An IBR operator is robust in the sense that on average it performs best over the whole uncertainty class.
  - The prior distribution reflects our existing knowledge.

Optimal Bayesian Operator

- Suppose that, in addition to a prior distribution coming from existing knowledge, there is a data sample *S*. The prior distribution conditioned on the sample yields a *posterior distribution* $\pi^*(\theta) = \pi(\theta|S)$.
  - The posterior distribution is evaluated via the likelihood function of the observed values.
- An IBR operator for the posterior distribution is called an *optimal Bayesian operator*.

Optimal Classification

- Scientific model: feature-label distribution governing pairs ($\mathbf{X}$, Y), where $\mathbf{X} = (X_1,\ldots, X_k)$ and $Y = 0$ or $Y = 1$.
  - Class-conditional distributions $f(\mathbf{x}|0)$ and $f(\mathbf{x}|1)$ govern the feature vectors in class 0 and class 1, respectively, along with class probabilities c_0 and c_1.
- A *classifier* ψ is a decision function on the set of feature vectors: $\psi(\mathbf{X}) = 0$ or $\psi(\mathbf{X}) = 1$.
  - Error of ψ is the probability of erroneous classification.
- *Bayes classifier* has minimum error: $\psi_{\mathrm{Bay}}(\mathbf{x}) = 1$ if $f(\mathbf{x}|1) \geq f(\mathbf{x}|0)$, and $\psi_{\mathrm{Bay}}(\mathbf{x}) = 0$ otherwise.
  - Error of ψ_{Bay} is the Bayes *error*.

Texas A&M GSPLab

Optimal Bayesian Classifier

- For any $\psi \in \Psi$, $E_{\pi^*}[\varepsilon(\psi_{\mathrm{OBC}}, \theta)] \leq E_{\pi^*}[\varepsilon(\psi, \theta)]$.
- Representation via effective class-conditional density:

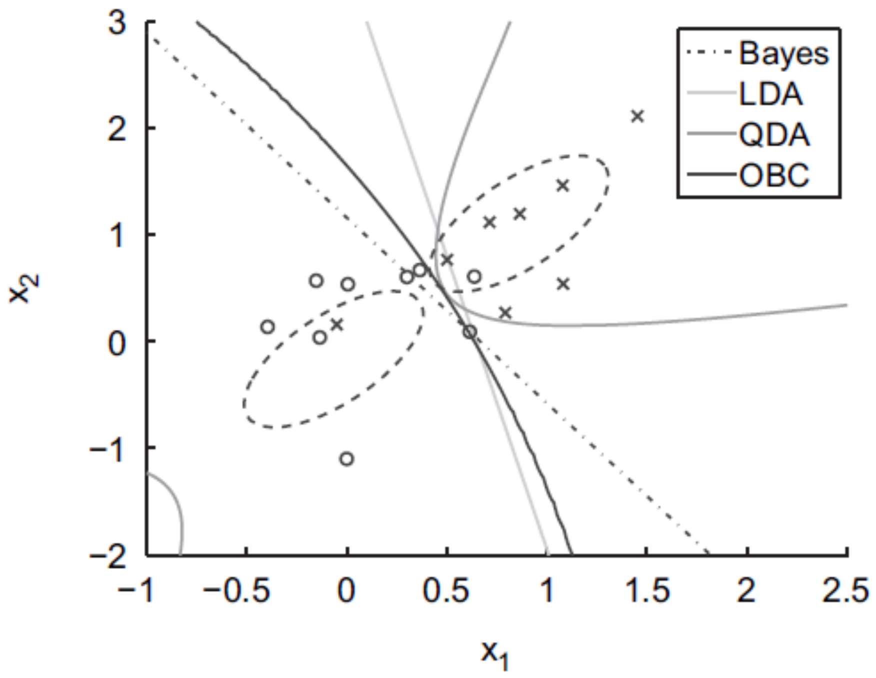
$$f(\mathbf{x}|y) = \int_{\Theta_y} f_{\theta_y}(\mathbf{x}|y)\,\pi^*(\theta_y)\,d\theta_y$$

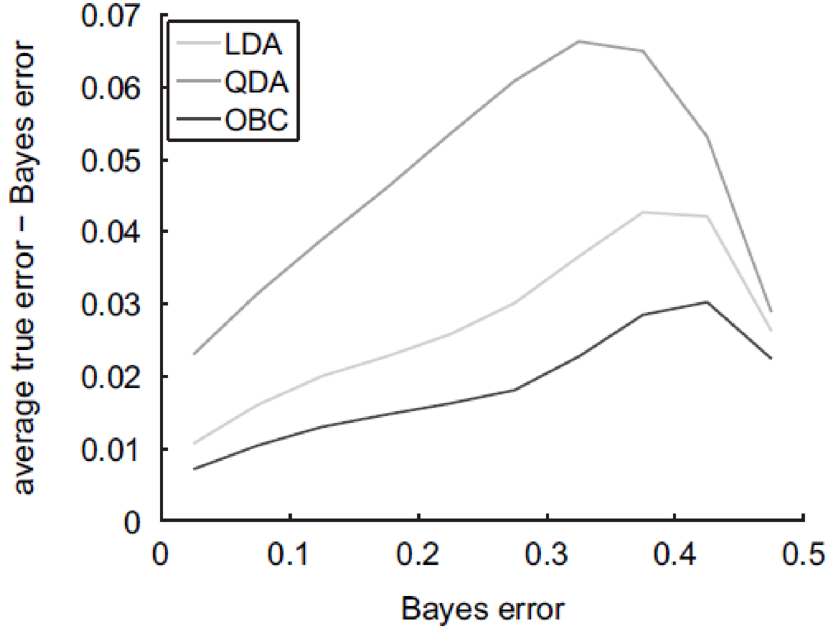
$$\psi_{\mathrm{OBC}}(\mathbf{x}) = \begin{cases} 0 & \text{if } \mathrm{E}_{\pi^*}[c] f(\mathbf{x}|0) \geq (1 - \mathrm{E}_{\pi^*}[c]) f(\mathbf{x}|1) \\ 1 & \text{otherwise.} \end{cases}$$

  - Dalton, L., and E. R. Dougherty, "**Optimal Classifiers with Minimum Expected Error within a Bayesian Framework – Parts I and II**," *Pattern Recognition*, **46**(5), 1288-1314, 2013.

Common-covariance Gaussian Model

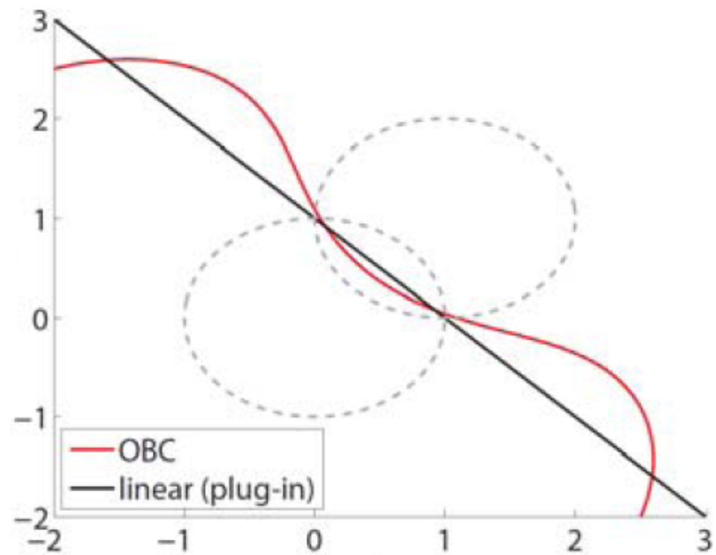
- Bayes classifier for actual model is linear and OBC is quadratic (n = 18).





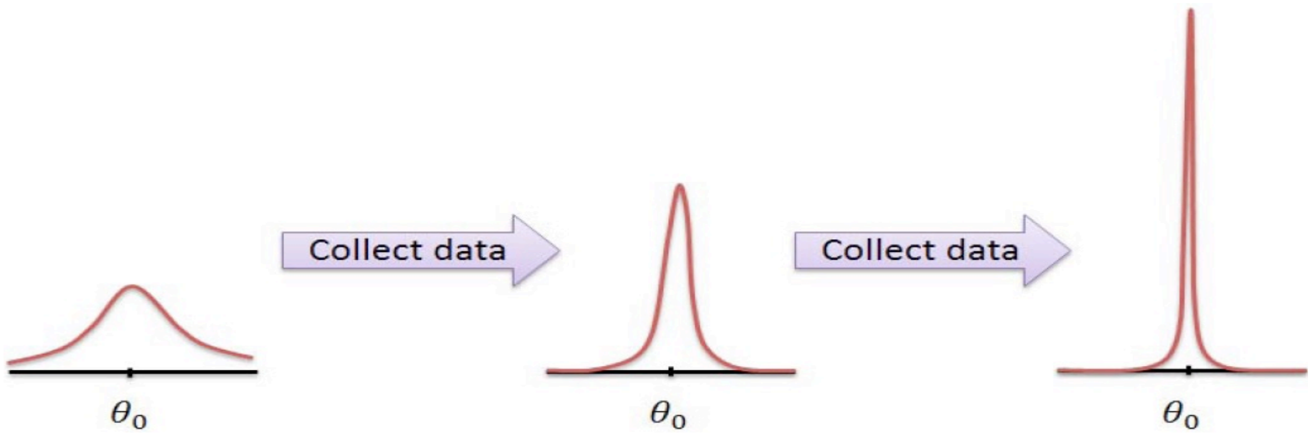
OBC for Gaussian Model

- Polynomial Optimal Bayesian Classifier (red line)
  - Dotted lines are level curves for the Gaussian class-conditional densities corresponding to the expected means and (equal) covariances for a given posterior.
  - Black solid line is linear classifier corresponding to the Bayes classifier for the expected mean and covariance parameters.



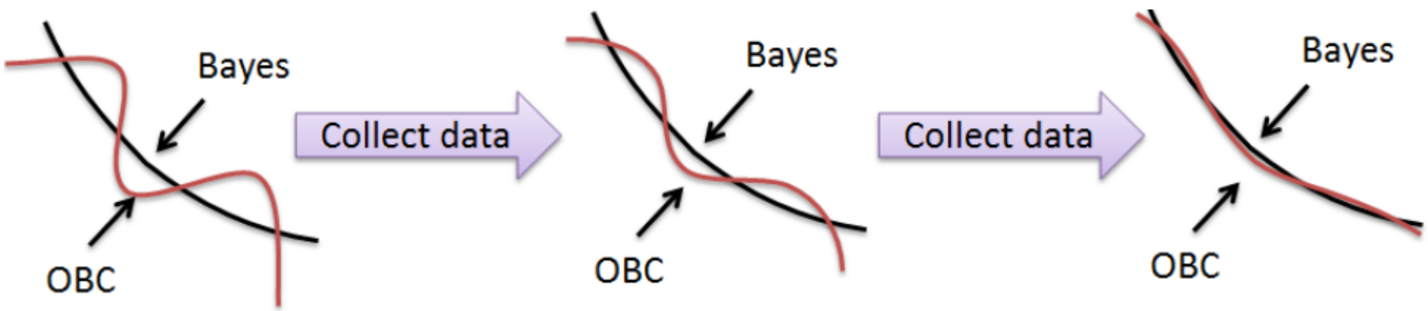
Prior + Data = Posterior

- The prior distribution represents the state of our knowledge prior to the data; the posterior represents the state of our knowledge after joining the prior with the data.
  - Data reduces the uncertainty in the prior – less variance.
  - Also centers posterior on the true model.



Convergence of the OBC

- In the Gaussian and multinomial models, as $n \to \infty$ the OBC converges to the Bayes classifier for the true feature-label distribution.



Texas A&M GSPLab

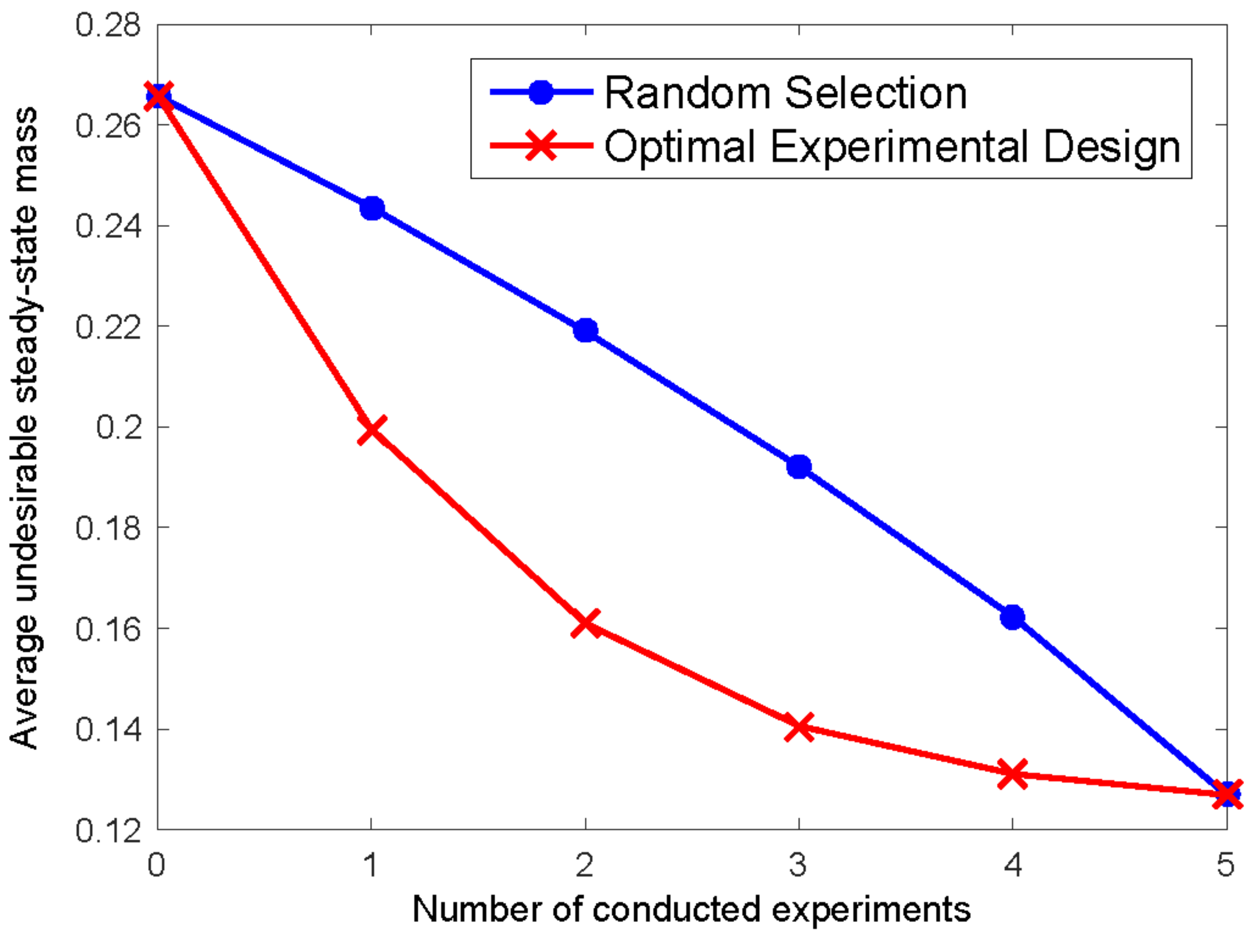
Objective Cost of Uncertainty

- An IBR operator is optimal over the uncertainty class, but it is likely to be suboptimal relative to the true model.
- There is a cost relative to applying ψ_θ on θ because $C_\theta(\psi_\theta) \leq C_\theta(\psi_{\mathrm{IBR}})$.
- For any $\theta \in \Theta$, the *objective cost of uncertainty* (OCU) is $\mathrm{OCU}(\theta) = C_\theta(\psi_{\mathrm{IBR}}) - C_\theta(\psi_\theta)$.
- *Mean objective cost of uncertainty*, $\mathrm{MOCU}(\Theta) = E_\Theta[\mathrm{OCU}(\theta)]$.
  - MOCU is objective-based quantification of uncertainty.
  - Entropy provides uncertainty quantification relative to the prior (posterior) distribution, not the engineering objective.

Objective-based Experimental Design

- If MOCU ≈ 0, then, on average, $C_\theta(\psi_{IBR}) \approx C_\theta(\psi_\theta)$.
  - If prior is concentrated around full model (plus some regularity conditions), expect IBR to be close to optimal.
- To get a new posterior, choose experiment with minimum expected MOCU given the experiment.
  - For each possible experiment, compute MOCU for all possible outcomes, average these MOCUs, take the minimum of these averages, and do experiment.
  - Proceed iteratively.
- Result *optimal experimental design* relative to the cost function and the objective uncertainty.
  - Maximal entropy reduction does not achieve this.

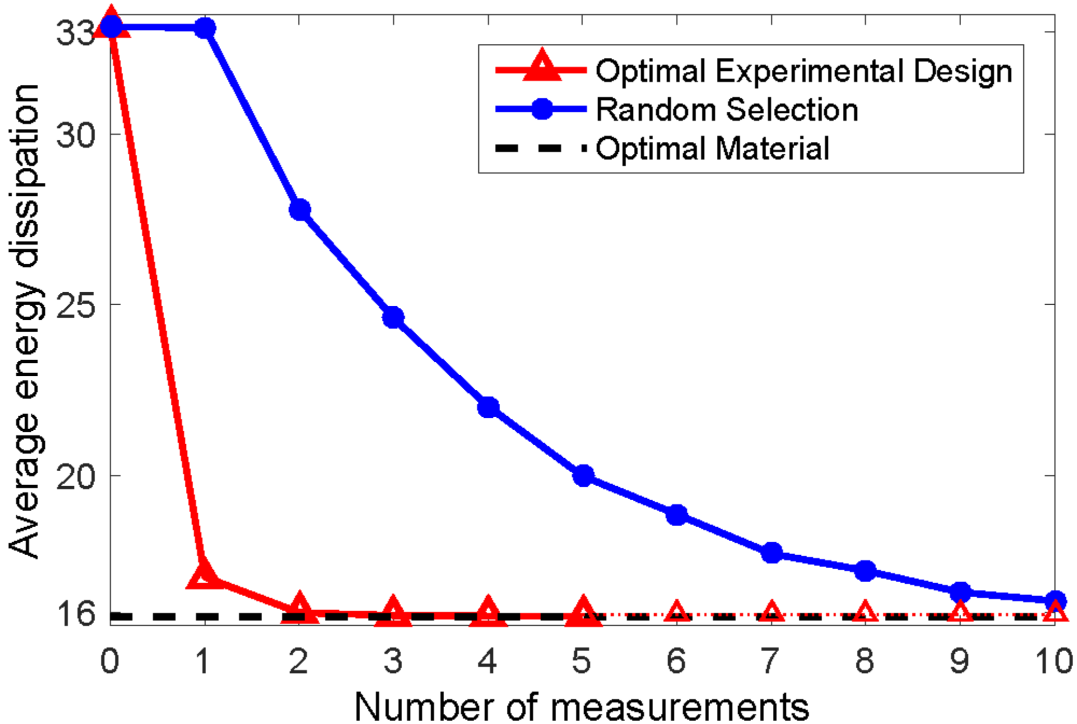
Design Drug Intervention in Gene Network



Texas A&M GSPLab

Design Shape Memory Alloy

- Design a shape memory alloy with low dissipation energy.
  - Aim of the experimental design is to suggest the best dopant and concentration for the next measurement.

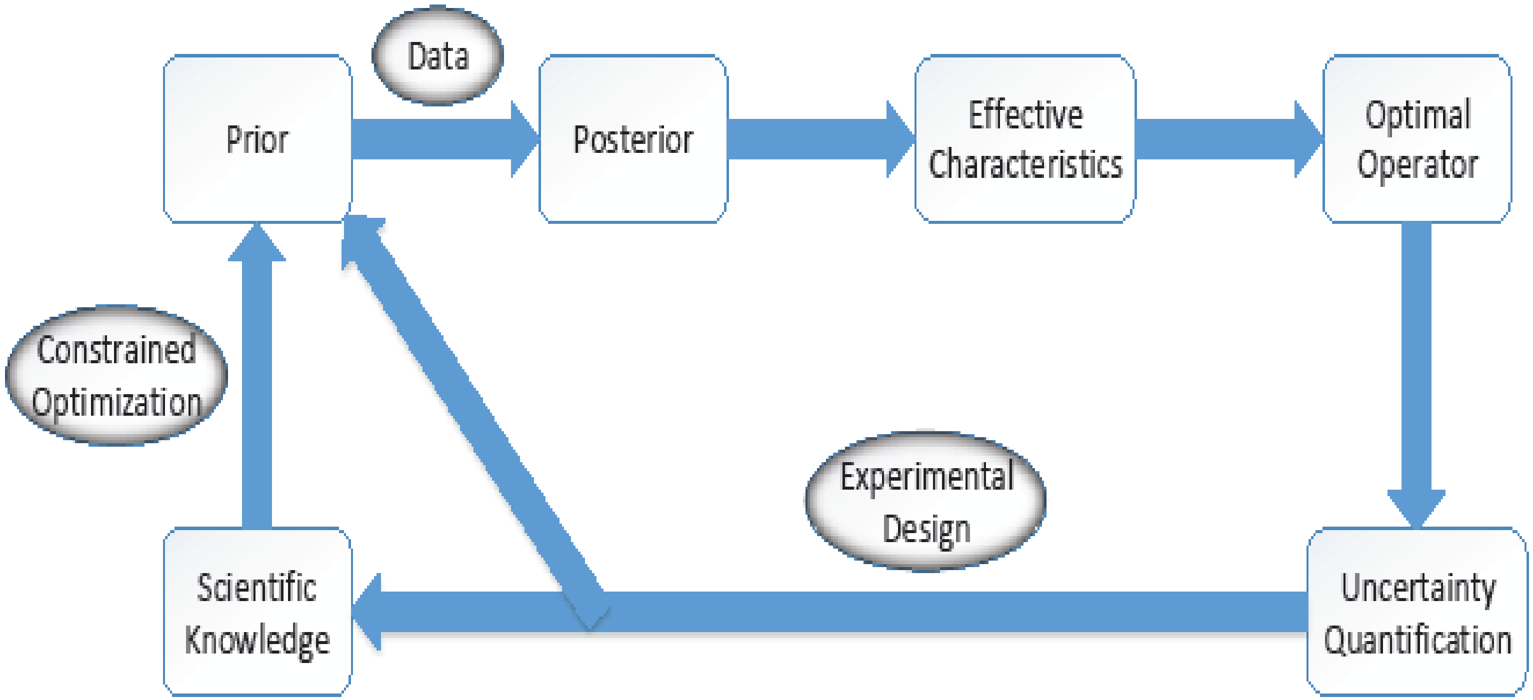


Prior Construction

- Base of the design loop. Numerous methods; however, a very general procedure can be used to derive the *Maximal Knowledge-driven Information Prior* (MKDIP) that minimizes an information-based cost function subject to constraints characterizing our prior knowledge.
- $\arg\min_{\pi \in \Pi} E_\pi[C_\theta(\xi, D)]$
  - $C_\theta(\xi, D)$ is a cost function depending on θ, the state ξ of our prior knowledge and part of the sample data D.
  - Maximum Entropy, Maximal Data Information, Expected Mean Log-Likelihood.
- MKDIP with constraints: optimization subject to constraints based on prior knowledge.

Texas A&M GSPLab

Design Loop



Design-Loop Operations

- Construct posterior from existing scientific knowledge.
- Update prior to posterior using data.
- Find effective characteristics (effective class-conditional densities).
- Find optimal Bayesian operator (OBC).
- Evaluate MOCU.
- Optimal experiment produces new knowledge to add to original knowledge or directly condition original prior, thus giving a new prior to re-institute the design process.
- Two optimizations in design loop: two cost functions, one for prior construction and one for operator design.

Conclusion

- Optimal operator design is the fundamental problem of engineering, whose aim is to design operators to alter the behavior of a physical systems in a desired manner.
- Optimal Bayesian operator design is natural: the operator is optimal relative to both our engineering objective and the state of our knowledge.
  - Scientific uncertainty is modeled, not operator uncertainty.
- If a problem is simple and there is a very large amount of data, then one can posit a complex operator form and estimate its parameters to hopefully get close to optimality.
  - But how close, if one does not have an optimal operator?
- Experiments should reduce pertinent scientific uncertainty.