An efficient optimization framework for gray-box conservation law simulation

Thesis proposal defense

PhD candidate: Han Chen

Committee: Qiqi Wang (chair), Karen Willcox, Youssef Marzouk
External evaluator: Hector Klie†

Massachusetts Institute of Technology
† ConocoPhillips Inc.

May 5, 2015

Outline

- Background.
- Thesis objective.

- Estimate gradient by twin model.
- Optimization framework.
- Application to turbulent flow optimization.

- Expected contribution.
- Proposed schedule.

Outline

- Background.
- Thesis objective.

- Estimate gradient by twin model.
- Optimization framework.
- Application to turbulent flow optimization.

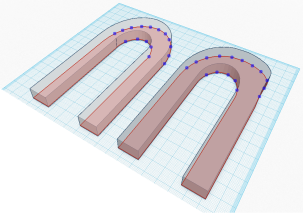
- Expected contribution.
- Proposed schedule.

Research background

- Interested in optimization constraint by conservation law simulation.
- Conservation law simulation can be expensive.
- The design space can be high-dimensional.
- Efficient sensitivity (adjoint) analysis may not be available.

A forest of oil wells in California, 1937

Internal cooling of turbine airfoil
Source: http://www.amaterastyo.biz/eng/technologies.html



Design space can be high-dimensional.

Research scope

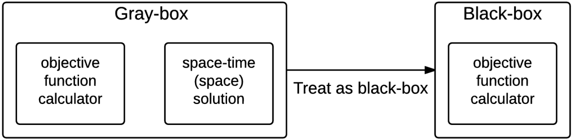
- Gray-box conservation law simulation:
  - adjoint not available.
  - governing PDE and its implementation not available.
  - can output space(-time) solutions.

	PDE and implementation	space(-time) solution	Adjoint
Black-box	✘	✘	✘
Gray-box	✘	✔	✘
Open-box	✔		✔

- High-dimensional design space.
  - large number of parameters required to parameterize the space(-time) dependent design.

Review of optimization methods

- If black-box, use derivative-free optimization,
  (pattern search methods [Tarma03], evolution based methods[Eberhart 10, Davis 10])
  - not require derivative evaluation.
  - not suitable for high-dimension optimization.
- If open-box, use gradient-based optimization,
  (quasi-Newton methods [John 77]: BFGS, L-BFGS, etc.)
  - requires efficient gradient evaluation, generally using adjoint.
  - suitable for high-dimension optimization.
- If gray-box,



Research objective

- Develop a method to estimate the gradient by using the space(-time) solution from the gray-box simulation.
- Formulate an optimization framework that uses the estimated gradient for efficient high-dimensional optimization.
- Given a fixed computational budget, assess how much design objective improvement can be achieved by using the proposed framework.

Outline

- Background.
- Thesis objective.

- Estimate gradient by twin model. (objective 1)
- Optimization framework. (objective 2)
- Application to turbulent flow optimization. (objective 3)

- Expected contribution.
- Proposed schedule.

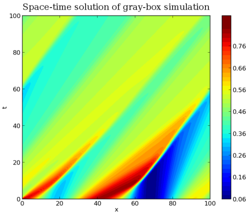
Leverage the space(-time) solution

Estimate the gradient:

- infer the conservation law from the space(-time) solution.
- apply adjoint to estimate gradient.

Example: infer flux $F(u)$ from space-time solution.

$$\frac{\partial u}{\partial t} + \frac{\partial F(u)}{\partial x} = c(x)$$
$$u(t=0,x) = u_0(x)$$
$$u(t,x=0) = u(t,x=1)$$



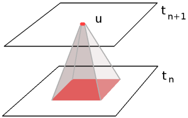
Propose to infer the flux or the source term that reproduce the space(-time) solution. The inferred conservation law is called twin model.

Why is the PDE inferrable?

- The governing PDE is a conservation law. Flux and source terms are functionals.

$$\times \quad \dot{u} = \mathcal{L}(u, c) \quad \checkmark \quad \frac{\partial u}{\partial t} + \frac{\partial F(u)}{\partial x} = c$$

- The flow quantities only depend on the flow quantities in an older time inside a domain of dependence.



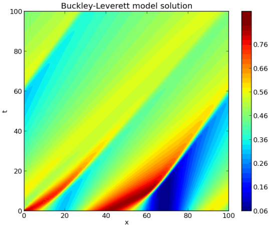
- Space-(time) solution can provide large number of samples.
- The inference can be independent of the design space dimensionality.

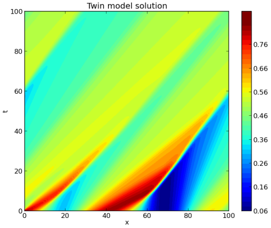
Minimize space(-time) solution mismatch

The inference can boil down to an optimization problem.

$$\frac{\partial u}{\partial t} + \frac{\partial F(u)}{\partial x} = c(x)$$
$$u(t=0,x) = u_0(x)$$
$$u(t,x=0) = u(t,x=1)$$

$$\frac{\partial \tilde{u}}{\partial t} + \frac{\partial \tilde{F}(\tilde{u})}{\partial x} = c(x)$$
$$\tilde{u}(t=0,x) = u_0(x)$$
$$\tilde{u}(t,x=0) = \tilde{u}(t,x=1)$$





$$\min_{\tilde{F}} \left\{ L(\tilde{F}) \equiv \int_t \int_x \|u - \tilde{u}\| \; dt\, dx \right\},$$

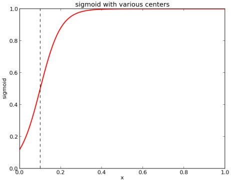
where $\|\cdot\|$ is a norm to be chosen.

Some technical details

- Parameterize the flux or source term
  polynomial, Fourier, wavelet, etc
  - we choose a family of sigmoid functions with various centers.
    $\tilde{F}(\cdot) = \sum_{i=1}^{n} \xi_i s(\cdot)$
- Basis selection
  matching pursuit [Adler 96, Billing07]: forward selection, backward pruning;
  regularization [Stone 77, Schwarz 78, Tibshirani 96]: AIC/BIC, Lasso, elastic net
  - we choose Lasso regularization for basis selection.
    $\min_{\tilde{F}} \left\{ L(\tilde{F}) + \lambda \sum_{i=1}^{n} |\xi_i| \right\}$

Some technical details

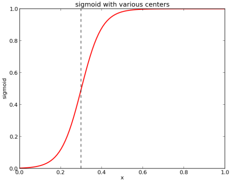
- Parameterize the flux or source term
  polynomial, Fourier, wavelet, etc
  - we choose a family of sigmoid functions with various centers.
    $\tilde{F}(\cdot) = \sum_{i=1}^{n} \xi_i s(\cdot)$



- Basis selection
  matching pursuit [Adler 96, Billing07]: forward selection, backward pruning;
  regularization [Stone 77, Schwarz 78, Tibshirani 96]: AIC/BIC, Lasso, elastic net
  - we choose Lasso regularization for basis selection.
    $\min_{\tilde{F}} \left\{ L(\tilde{F}) + \lambda \sum_{i=1}^{n} |\xi_i| \right\}$

Some technical details

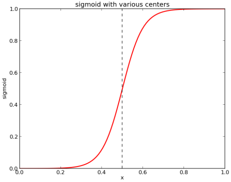
- Parameterize the flux or source term
  polynomial, Fourier, wavelet, etc
  - we choose a family of sigmoid functions with various centers.
    $\tilde{F}(\cdot) = \sum_{i=1}^{n} \xi_i s(\cdot)$



- Basis selection
  matching pursuit [Adler 96, Billing07]: forward selection, backward pruning;
  regularization [Stone 77, Schwarz 78, Tibshirani 96]: AIC/BIC, Lasso, elastic net
  - we choose Lasso regularization for basis selection.
    $\min_{\tilde{F}} \left\{ L(\tilde{F}) + \lambda \sum_{i=1}^{n} |\xi_i| \right\}$

Some technical details

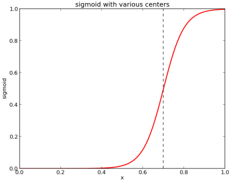
- Parameterize the flux or source term
  polynomial, Fourier, wavelet, etc
  - we choose a family of sigmoid functions with various centers.
    $\tilde{F}(\cdot) = \sum_{i=1}^{n} \xi_i s(\cdot)$



- Basis selection
  matching pursuit [Adler 96, Billing07]: forward selection, backward pruning;
  regularization [Stone 77, Schwarz 78, Tibshirani 96]: AIC/BIC, Lasso, elastic net
  - we choose Lasso regularization for basis selection.
    $\min_{\tilde{F}} \left\{ L(\tilde{F}) + \lambda \sum_{i=1}^{n} |\xi_i| \right\}$

Some technical details

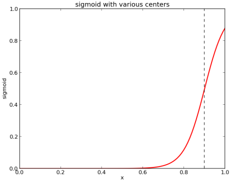
- Parameterize the flux or source term
  polynomial, Fourier, wavelet, etc
  - we choose a family of sigmoid functions with various centers.
    $\tilde{F}(\cdot) = \sum_{i=1}^{n} \xi_i s(\cdot)$



- Basis selection
  matching pursuit [Adler 96, Billing07]: forward selection, backward pruning;
  regularization [Stone 77, Schwarz 78, Tibshirani 96]: AIC/BIC, Lasso, elastic net
  - we choose Lasso regularization for basis selection.
    $\min_{\tilde{F}} \left\{ L(\tilde{F}) + \lambda \sum_{i=1}^{n} |\xi_i| \right\}$

Some technical details

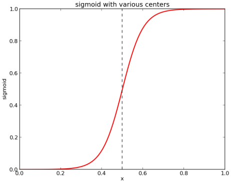
- Parameterize the flux or source term
  polynomial, Fourier, wavelet, etc
  - we choose a family of sigmoid functions with various centers.
    $\tilde{F}(\cdot) = \sum_{i=1}^{n} \xi_i s(\cdot)$



- Basis selection
  matching pursuit [Adler 96, Billing07]: forward selection, backward pruning;
  regularization [Stone 77, Schwarz 78, Tibshirani 96]: AIC/BIC, Lasso, elastic net
  - we choose Lasso regularization for basis selection.
    $\min_{\tilde{F}} \left\{ L(\tilde{F}) + \lambda \sum_{i=1}^{n} |\xi_i| \right\}$

Some technical details

- Parameterize the flux or source term
  polynomial, Fourier, wavelet, etc
  - we choose a family of sigmoid functions with various centers.
    $\tilde{F}(\cdot) = \sum_{i=1}^{n} \xi_i s(\cdot)$



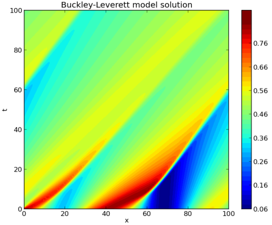
- Basis selection
  matching pursuit [Adler 96, Billing07]: forward selection, backward pruning;
  regularization [Stone 77, Schwarz 78, Tibshirani 96]: AIC/BIC, Lasso, elastic net
  - we choose Lasso regularization for basis selection.
    $\min_{\tilde{F}} \left\{ L(\tilde{F}) + \lambda \sum_{i=1}^{n} |\xi_i| \right\}$

A demonstration of twin model

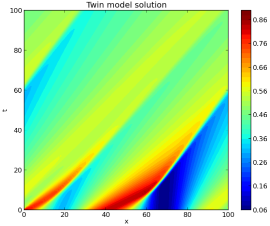
Consider an objective: [Kucuk 06]

$$\min_{c \in \mathbb{R}} \left\{ J(u) \equiv \int_{x=0}^{1} \left| u(x, t=1; c) - u^*(x) \right|^2 \mathrm{d}x \right\}$$

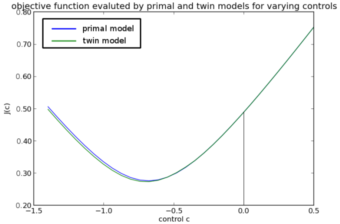
constrained by Buckley-Leverett flow, where c is a constant source term to be optimized. u^* is a given spatial profile.



Gray-box model solution



Twin model solution



J calculated by the two models

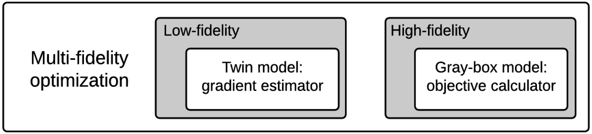
Outline

- Background.
- Thesis objective.

- Estimate gradient by twin model. (objective 1)
- Optimization framework. (objective 2)
- Application to turbulent flow optimization. (objective 3)

- Expected contribution.
- Proposed schedule.

Review of multi-fidelity optimization (MFO)



- MFO methods include
  - pattern search MFO [Booker 99].
  - trust-region MFO [Wild 13, March 12, Robinson 06].
  - Bayesian MFO [Kennedy 01, March 11].
- Choose Bayesian MFO as our optimization framework:
  - uses all high-fidelity model evaluation to find the next design.
  - can fuse sampled data of different types: co-Kriging [Chung 02].
  - the next candidate design is optimal under a Bayesian metric.

Model objective and gradient Bayesianly

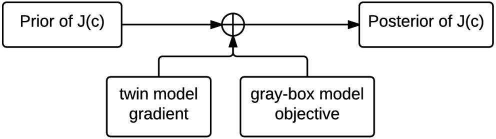
- Gaussian process modeling:
  - $J(c)$: the objective calculated by the gray-box model.
  - $\epsilon(c)$: the error in the objective's gradient calculated by the twin model.
- Relate gray-box model's objective with twin model's gradient:

$$\begin{cases} g(c) = \nabla J(c) + \epsilon(c) \\ \operatorname{cov}\left[\nabla J(c_1), \epsilon(c_2)\right] = 0 \\ \operatorname{cov}\left[J(c_1), \epsilon(c_2)\right] = 0 \end{cases} \qquad \text{for any } c, c_1, c_2\,,$$

where $g(c)$ is the objective's gradient calculated by the twin model.

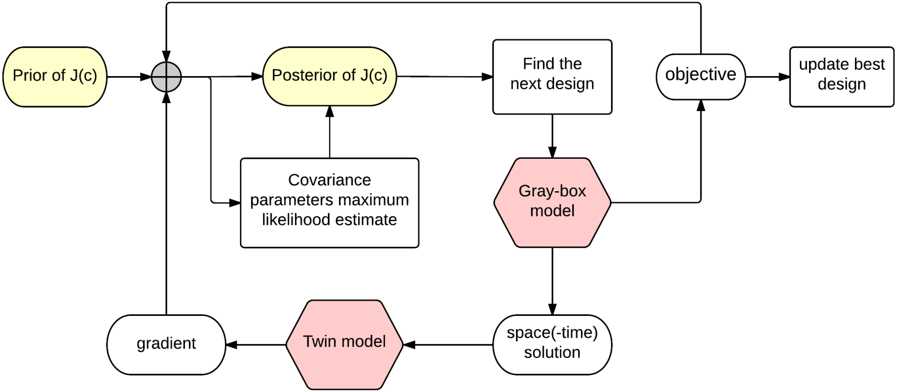
Perform Bayesian optimization

- Update $J(c)$ Bayesianly



- Find the next design to evaluate the gray-box model [Snoek 12]
  - Define “improvement”: $\max\{J(c_{best}) - J(c), 0\}$.
  - Choose the next design as the maximizer of the expected improvement.

Optimization framework

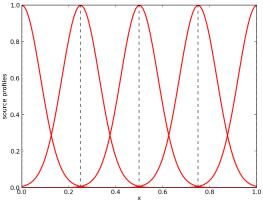


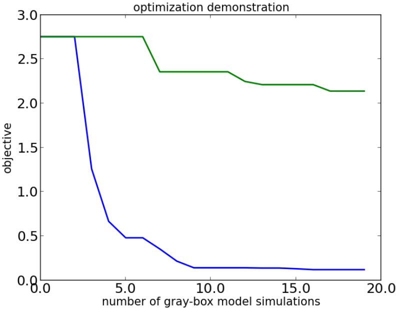
A demonstration of optimization

- Source parameterized by 5 design variables.
- Optimization problem:

$$\min_{c \in \mathbb{R}^5} \left\{ J(u) \equiv \int_{x=0}^{1} \left| u(x, t=1; c) - u^*(x) \right|^2 \mathrm{d}x + \lambda \sum_{i=1}^{5} c_i^2 \right\},$$

where $u^*(x) = u^*(x, t=1; c^*)$, u^* generated by gray-box model.





github.com/septfleur/twinmodel.git

Outline

- Background.
- Thesis objective.

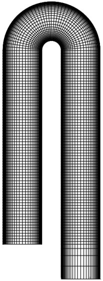
- Estimate gradient by twin model. (objective 1)
- Optimization framework. (objective 2)
- Application to turbulent flow optimization. (objective 3)

- Expected contribution.
- Proposed schedule.

Optimize return bend geometry

- Optimize the geometry of an internal cooling hole in turbine airfoil to minimize the time averaged pressure loss. [Coletti 13]
- Model as a 2-D return bend problem.
- Design space can be high-dimension.

Internal cooling of turbine airfoil
Source: http://www.amaterastyo.biz/eng/technologies.html



Apply twin model optimization framework

- Flow is turbulent and incompressible, $Re \sim 40,000$, $Mach \sim 0.05$
- Candidate simulation models:
  - Time averaged quantities:

    RANS models: Reynolds stress models, eddy viscosity models (e.g. mixing length models, $k - \omega$ models), etc [Wilcox 98]
  - Space-time dependent quantities:

    LES, DNS, etc
- Apply twin model optimization framework:
  - Gray-box model: time averaged quantities of LES simulation.
  - Twin model: a RANS model with adaptive eddy viscosity modelling.

Outline

- Background.
- Thesis objective.

- Estimate gradient by twin model. (objective 1)
- Optimization framework. (objective 2)
- Application to turbulent flow optimization. (objective 3)

- Expected contribution.
- Proposed schedule.

Expected contributions

- Develop an efficient method to estimate gradients when the governing PDE is not available.
- Provide an optimization framework based on gray-box conservation law simulation for high-dimensional design problems.
- A demonstration of the twin model optimization framework in a high-dimensional turbulent flow optimization, showing superior objective function improvement given a fixed computational budget.

Proposed schedule

- Completed
  - Course work.
  - Formulation of twin model and its inference.
  - Development of twin model optimization framework.
  - Demonstration of optimization on a 1-D flow testcase.
- To be completed
  - May 15': Setup an 2-D LES solver for the return bend testcase in OpenFoam.
  - Jun 15': Setup a RANS solver with adaptive eddy viscosity in python.
  - Jul-Oct 15': Optimize return bend geometry.
  - Aug 15': Hold a committee meeting to report progress.
  - Sep-Nov 15': Write thesis.
  - Jan 16': Defense.

Thank you!

References I

RC. Eberhart et al.
Particle swarm optimization.
Encyclopedia of machine learning, 2010.

L. Davis
Handbook of genetic algorithms.
New York: Van Nostrand Reinhold, 1991.

K. Tamara et al.
Optimization by direct search: new perspective on some classical and modern methods.
SIAM review, 45.3:385-482, 2003.

D. Dennis et al.
Quasi-Newton methods, motivation and theory.
SIAM review, 19.1:46-89, 1977.

References II

R. Adler
Comparison of basis selection methods.
Signals, systems and computers, Thirtieth Asilomar Conference on. Vol. 1. IEEE, 1996.

S. Billing et al.
Feature subset selection and ranking for data dimensionality reduction.
Pattern analysis and machine intelligence, IEEE transactions on 29.1 (2007): 162-166.

G. Schwarz
Estimating the dimension of a model.
The annals of statistics 6.2 (1978): 461-464.

References III

M. Stone
An asymptotic equivalence of choice of model by cross-validation and Akaike's criterion.
Journal of the royal statistical society. series B (methodological) (1977): 44-47.

R. Tibshirani
Regression shrinkage and selection via the lasso.
Journal of the royal statistical society. series B (methodological) (1996): 267-288.

I. Kucuk
An efficient computational method for the optimal control problem for the Burgers equation.
Mathematical and computer modelling 44.11 (2006): 973-982.

References IV

A. Booker
A rigorous framework for optimization of expensive functions by surrogates.
Structural optimization 17.1 (1999): 1-13.

S. Wild et al.
Global convergence of radial basis function trust-region algorithms for derivative-free optimization.
SIAM Review 55.2 (2013): 349-371.

A. March
Multifidelity methods for multidisciplinary system design
Dissertation, Massachusetts Institute of Technology (2012)

References V

T. Robinson
Multifidelity optimization for variablecomplexity design.
Proceedings of the 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Portsmouth, VA. 2006.

M. Kennedy et al.
Bayesian calibration of computer models
Journal of the royal statistical society, series B (statistical methodology) 63.3 (2001): 425-464.

A. March et al.
Gradient-based multifidelity optimisation for aircraft design using Bayesian model calibration
Aeronautical Journal, 115.1174 (2011): 729.

References VI

- H. Chung et al.
  Using gradients to construct cokriging approximation models for high-dimensional design optimization problems.
  AIAA paper 317 (2002): 14-17.
- J. Snoek et al.
  Practical Bayesian optimization of machine learning algorithms.
  Advances in neural information processing systems, 2012.
- F. Coletti et al.
  Optimization of a U-Bend for Minimal Pressure Loss in Internal Cooling Channels-Part II: Experimental Validation.
  Journal of Turbomachinery 135.5 (2013): 051016.
- D. Wilcox
  Turbulence modeling for CFD.
  Vol. 2. La Canada, CA: DCW industries, (1998)

References VII

J. Gardner
Bayesian optimization with inequality constraints
Proceedings of The 31st International Conference on Machine Learning. 2014.

Discussion of optimization convergence 1

Theorem

The design sequence is dense.
Assume

- $\mathcal{X} \in \mathbb{R}^n$, $\|\cdot\|$ *be the* L_2 *norm defined on* $\mathcal{X}$.
- $\mathcal{H}$ *and* $\mathcal{H}'$ *be two reproducing kernel Hilbert spaces of functions on* $\mathcal{X}$, *with kernels* $K(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ *and* $K'(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \mapsto \mathbb{R}$ *respectively.*
- *There exist* $k : \mathbb{R}^+ \bigcup \{0\} \mapsto \mathbb{R}$ *and* $k' : \mathbb{R}^+ \bigcup \{0\} \mapsto \mathbb{R}$, *such that* K *and* K' *satisfies* $K(x,y) = k(\|x-y\|)$ *and* $K'(x,y) = k'(\|x-y\|)$ *respectively, for* $\forall x, y \in \mathcal{X}$.
- k *and* k' *has the Fourier transforms* $\hat{k}$ *and* $\hat{k}'$ *respectively. They satisfy the asymptotic properties* $\hat{k}(u) = \Theta(|u|^{-n-2\nu})$ *and* $\hat{k}'(u) = \Theta(|u|^{-n-2\nu'})$, *as* $|u| \to \infty$, *with* $\frac{1}{2} < \nu < \infty$ *and* $\nu' = \nu - 1$. *(* Θ *is the asymptotic big* Θ *notation.)*

Then all functions in $\mathcal{H}$ *are differentiable.*
In addition, let

- $f \in \mathcal{H}$.
- $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^T$, $\epsilon_i \in \mathcal{H}'$, $i = 1, \cdots, n$. ϵ_i *is independent of* ϵ_j *for* $i \neq j$.
- $g = \epsilon + \nabla f$.

Suppose an infinitely long sequence is generated by the following strategy:

$$x_{s+1} = \arg\max_{x \in \mathcal{X}} \mathbb{E}\left[\max\left(f(x) - f(x_s^*), 0\right) \middle| \mathcal{S}\right]$$

$$x_s^* = \arg\max_{x \in \{x_1, \cdots, x_s\}} f(x)$$

$$\mathcal{S} = \left\{ \{x_1, \cdots, x_s\}, \{f(x_1), \cdots, f(x_s)\}, \{g(x_1), \cdots, g(x_s)\} \right\}$$

Then the sequence $\{x_1, x_2, \cdots\}$ *is dense in* $\mathcal{X}$ *for* $\forall x_1 \in \mathcal{X}$.

Dicussion of optimization convergence 2

- Mostly interested in improving the objective given a limited computational budget.
- Gray-box simulation can be so expensive that only a small number of evaluations are feasible.
- When the design is close to the optimal, the optimization framework degenerates to DFO.
- A converngence proof helps complete the theory, but may offer little practical value in realistic problems.

Exploit twin model more than its gradient?

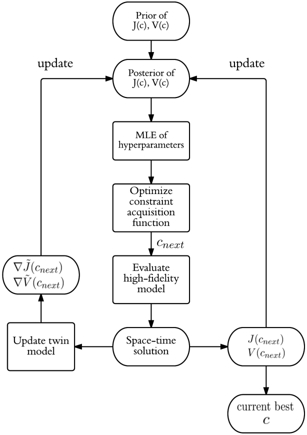
- Proposed framework collocates (in the design space)
  - gray-box objective evaluation.
  - twin model training.
  - twin model gradient evaluation.
- May benefit from non-collocated optimization framework, for example
  - perform a sub-optimization (gradient-based) using twin model in a trust-region.
- However, twin model is not a “dispensable” asset in terms of evaluation cost.

Extension to constraint optimization

- Categorize constraints:
  - constraint evaluation that requires PDE simulation.
  - constraint evaluation that not requires PDE simulation.
- Constraint expected improvement:

$$\mathbb{E}\left[\max\left(J(c) - J(c_s^*), 0\right) \middle| \mathcal{S}\right] \cdot \mathbb{P}\left[V(c) \leq 0 \middle| \mathcal{S}\right],$$

- Constraint twin model optimization framework.

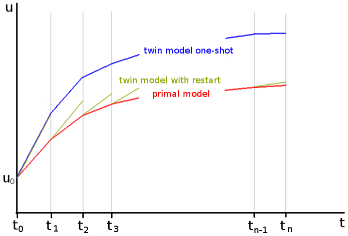


Basis selection without PDE solve

For time dependent twin model,

$$Err_G = \frac{1}{T}\sum_{i=1}^{N}\sum_{k=1}^{T}\left(\tilde{u}_{ik} - u_{ik}\right)^2 \Delta t_k \, |\Delta \mathbf{x}_i|$$

$$Err_L = \frac{1}{T}\sum_{i=1}^{N}\sum_{k=1}^{T}\left(\tilde{u}'_{ik} - u_{ik}\right)^2 \Delta t_k \, |\Delta \mathbf{x}_i|$$



Theorem

Global-local error

Consider the timestepwise mapping of the twin model

$$G:\ \mathbb{R}^n \mapsto \mathbb{R}^n,\ \tilde{u}^i \to G\tilde{u}^i = \tilde{u}^{i+1}\ , \quad i = 1, \cdots, n\,.$$

If G is a Lipschitz continuous mapping with constant α

$$\|Gx - Gy\|_{L_2} \le \alpha \|x - y\|_{L_2}$$

then

$$Err_G \le (1 + \alpha + \cdots + \alpha^{n-1}) Err_L\,.$$

If $\alpha < 1$*, then*

$$Err_G < \frac{1}{1-\alpha} Err_L$$

Major and minor programs of study

- Aerospace computational engineering
  - 2.25 advanced fluid dynamics
  - 16.920 numerical methods for PDE
  - 6.437 inference and information
  - statistical estimation and control
  - 18.335 introduction to numerical methods
  - 16.940 numerical methods for stochastic modeling
  - 6.438 algorithms for inference
- Finance
  - mas.665/15.375 development ventures
  - 15.450 analytics of finance
  - 15.s26