MDO, GP, RONWHAT ELSE?

Prof. Joseph Morlier, MLclass 2019





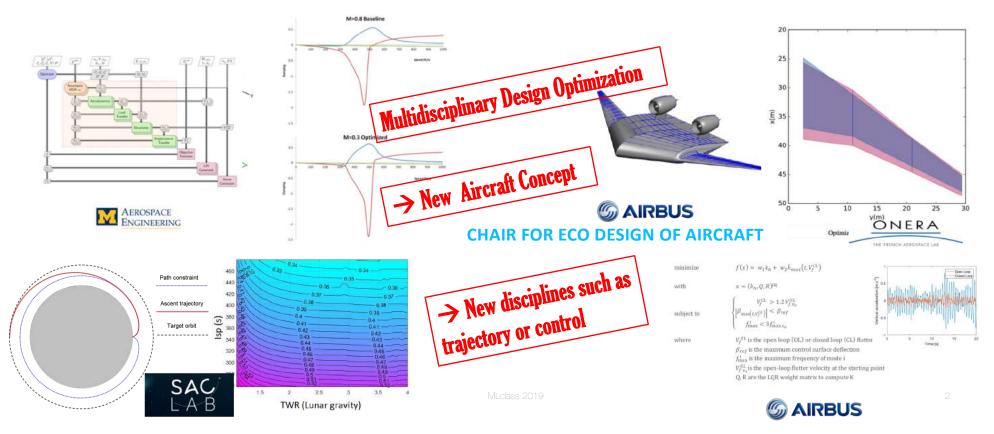




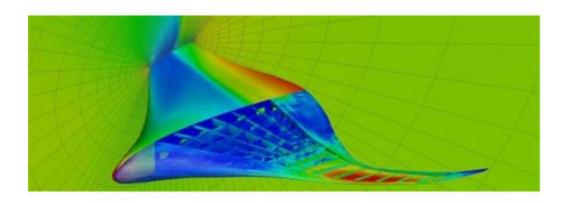


My Research Group (Joint research with ONERA on MDO)

• 4 PhDs, 1 postdoc, 1 research assistant, 4 MsCs



Popularization for our common research ONERA-SUPAERO



http://mdolab.engin.umich.edu

Optimization [MDO] for connecting people?

https://www.linkedin.com/pulse/optimiz ation-mdo-connecting-people-joseph-

ioseph morlier Professor in Structural and Multidisciplinary

Design Optimization, ... any idea?





Publié le 14 février 2019

2 articles

Outlines for today

multidisciplinary Design optimization

generates Data (a lot)

1. MDO

2. GP

3. ROM

multidisciplinary optimization

RECIPES

• MDO+GP+ROM = Optimization of coupled (costly) simulation codes at fixed budget

MOPTA08 is a <u>multidisciplinary design optimization</u> (MDO) benchmark problem based on a real-life problem from the automotive industry.

It states a large-scale multidisciplinary mass optimization of a vehicle in a crash test simulation. Real simulation (HPC) can optimistically compute about **60 points/day**. It was highly desirable to solve the optimization problem $in \le 1$ month (30 days).

For you:

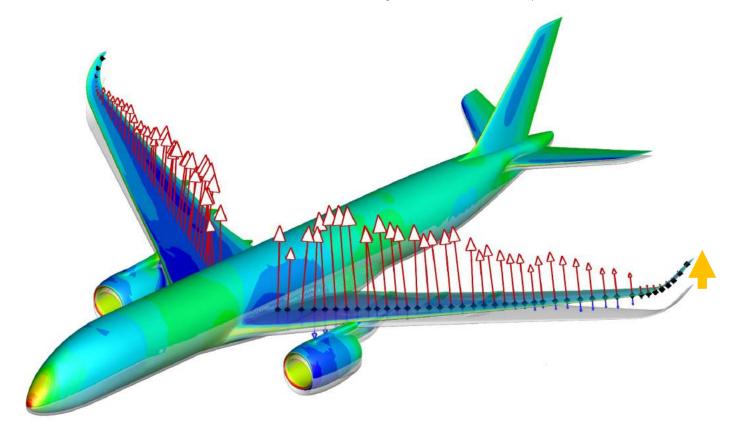
• 1 hour course (research oriented) 21/10 + 2H Python Practice 23/10

Outlines for today

2. GP

3. ROM

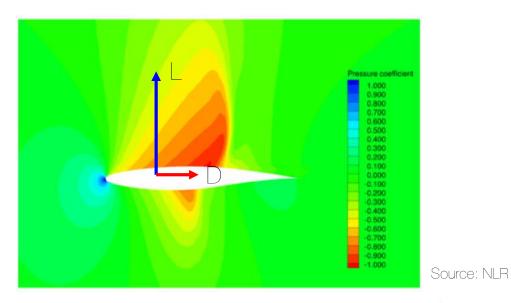
What is an MDA? Static Aeroelasticity for example?



Source: DLR

But first, what is Disciplinary Optimization?

Example: Aerodynamics (L/D max)

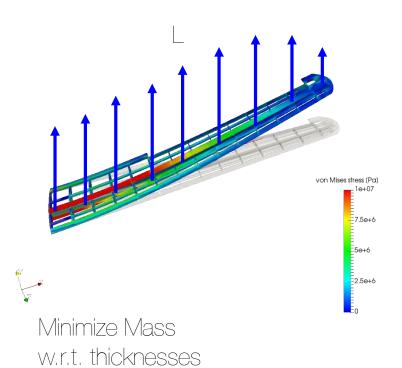


Minimize D w.r.t. shape, a

Subject to L = W

What is Disciplinary Optimization (2)?

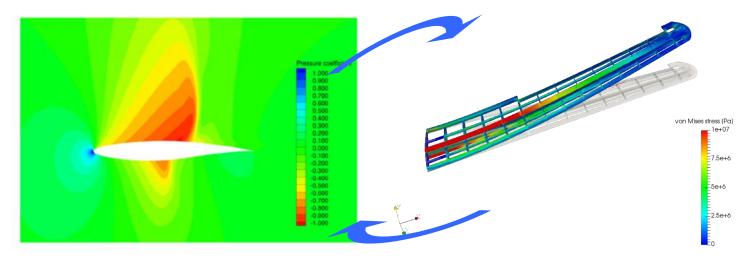
Another example: Structures



Source: simscale.com

Subject to $\sigma \leq \sigma y$

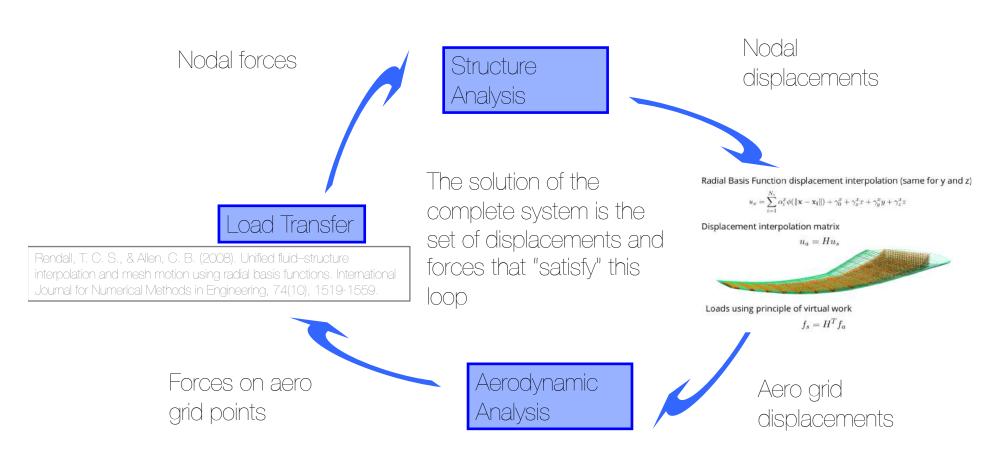
However... Disciplines are not isolated:



Structural deformation of wing > changes in the shape exposed to airflow

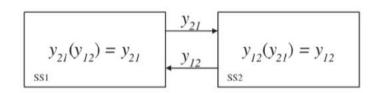
Changes in the shape exposed to airflow → changes in the aerodynamic loads

Then, how do we solve the complete system?



Multi-Disciplinary Analysis

- Computation of the state variables at equilibrium for given x and z
 - Generally computed using a fixed-point algorithm (Jacobi or Gauss-Seidel)
 - Or a root-finding method (Newton-Raphson)



(Step 0) choose initial guess y_{12}^0 , set i=0

(Step 1)
$$i = i + 1$$

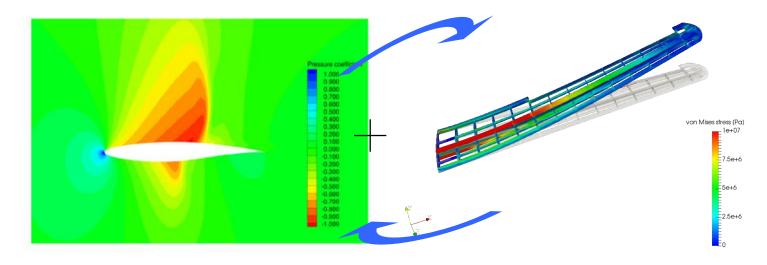
(Step 2)
$$y_{21}^i = y_{21}(y_{12}^{i-1})$$

(Step 3)
$$y_{12}^i = y_{12}(y_{21}^i)$$

(Step 4) if
$$|y_{12}^i - y_{12}^{i-1}| < \varepsilon$$
 stop, otherwise go to (Step 1)



→we need to analyze BOTH disciplines at the SAME TIME



Minimize D, or Mass, or a combination of D and Mass w.r.t. shape, a, thicknesses
Subject to:

 $L = \bigvee$

 $0 \leq 0$

In practice, how do we solve that problem?

One possible approach: MultiDisciplinary Feasible (MDF, probably the most intuitive one...)

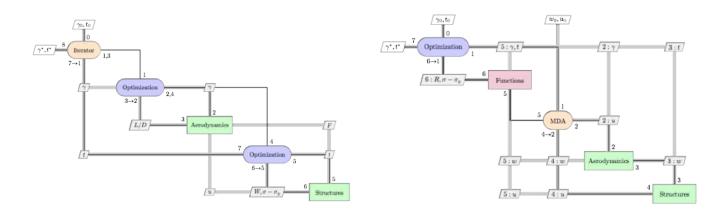
Steps:

- 1. Start from a set of particular design variables: shape, a, thicknesses
- 2. Solve the complete system (with all the interactions) for these values
- 3. Evaluate objective function and constraints
- 4. From these values, the optimizer proposes a new set of These steps are repeated until the optimum is reached. Next: MDO ... The big picture

MDO optimizes all variables simultaneously, accounting for all the couplings

Sequential optimization

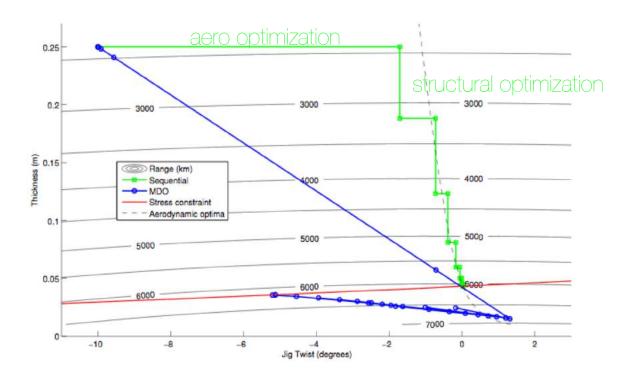
MDO



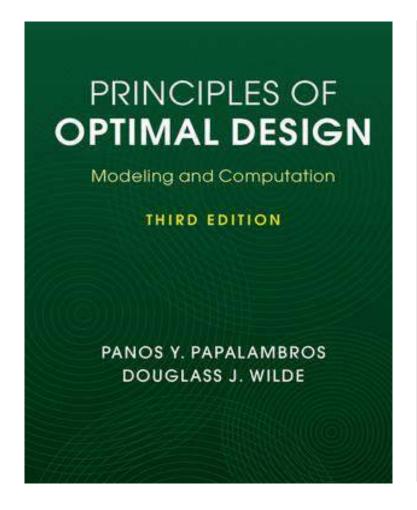
I. R. Chittick and J. R. R. A. Martins. An asymmetric suboptimization approach to aerostructural optimization. Optimization and Engineering, 10(1):133–152, Mar. 2009. doi:10.1007/s11081-008-9046-2.

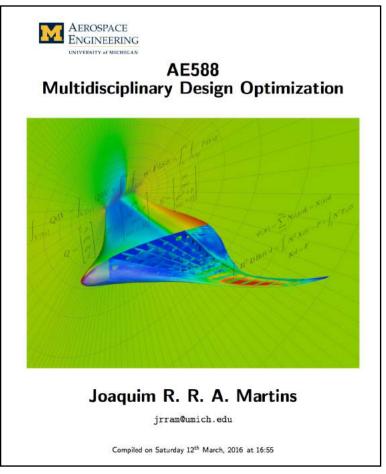
Sequential optimization fails to find the multidisciplinary optimum

Chittick, I. R., & Martins, J. R. (2008). Aero-structural optimization using adjoint coupled post-optimality sensitivities. Structural and Multidisciplinary Optimization, 36(1), 59-70.



Good Starting Point (x0)

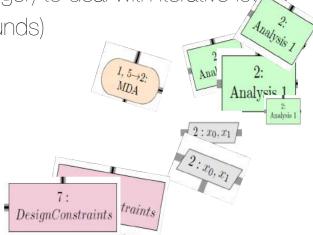




Assembling MDO systems

In order to assemble an MDO "architecture" we need a number of components: 0,7-1: Objective Function

- One (or more) optimizers
- One (or more) objectives
- A number of disciplinary tools (or disciplines, or competences)
- Possibly some coordinator (or converger) to deal with iterative logs,
- A bunch of design variables (with bounds)
- Some constraint specification



Assembling MDO systems

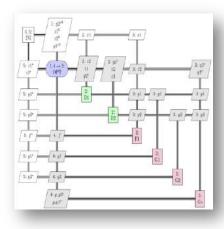
Multidisciplinary Design Optimization

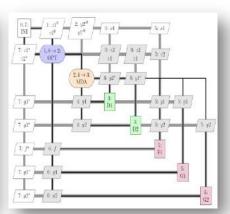
Monolithic

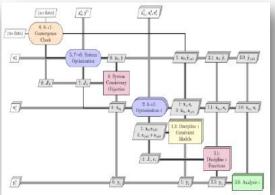
All-at-Once (AAO)
Simultaneous Analysis and Design (SAND)
Individual Discipline Feasible (IDF)
Multiple Discipline Feasible (MDF)

Distributed

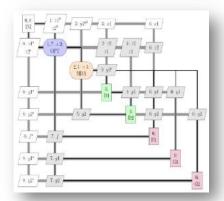
Concurrent Sub-Space Optimization (CSSO) Bi-Level System Synthesis (BLISS) Collaborative Optimization (CO) Analytical Target Cascading (ATC)

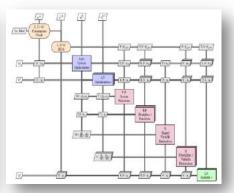


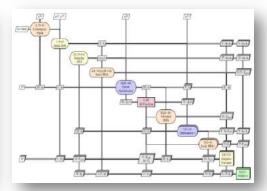




MDF Multidisciplinary Feasible approach—a complete analysis is performed at every optimization iteration. Also known as the All-in-One approach.







Illustrative example: the Sellar problem

2 disciplines involved Variables: X₁, y₁, y₂, Z₁, Z₂

We'll see later what are the differences between these variables ...

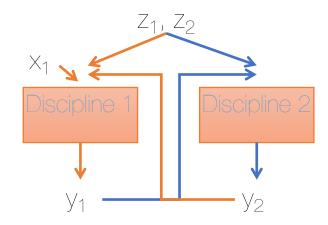
```
minimize x_1^2 + z_2 + y_1 + \exp(-y_2)
with respect to z, x or (z_1, z_2, x_1)
subject to:
3.16 - y_1 \le 0
y_2 - 24 \le 0
-10 \le z_1 \le 10
0 \le z_2 \le 10
0 \le x_1 \le 10
```

Discipline 1: $y_1(z_1, z_2, x_1, y_2) = z_1^2 + x_1 + z_2 - 0.2y_2$ Discipline 2: $y_2(z_1, z_2, y_2) = \sqrt{y_1} + z_1 + z_2$

ъенаг, н. ъ., ватн, ъ. М., and Renaud, J. E., "Response Surface Based, Concurrent Subspace Optimization for Multidisciplinary System Design", 34th Aerospace Sciences Meeting and Exhibit, Aerospace Sciences Meetings, 1996.

Illustrative example: the Sellar problem

- Design variables: z_1 , z_2 , x_1 to minimize the objective
- Shared (or global) variables: z₁, z₂
- Local variable: X₁
- Coupling variables: y₁, y₂



minimize
$$x_1^2 + z_2 + y1 + e^{-y_2}$$

with respect to z_1, z_2, x_1
subject to:
 $\frac{y_1}{3.16} - 1 \ge 0$
 $1 - \frac{y_2}{24} \ge 0$
 $-10 \le z_1 \le 10$
 $0 \le z_2 \le 10$
 $0 \le x_1 \le 10$

Discipline 1:
$$y_1(z_1, z_2, x_1, y_2) = z_1^2 + x_1 + z_2 - 0.2y_2$$

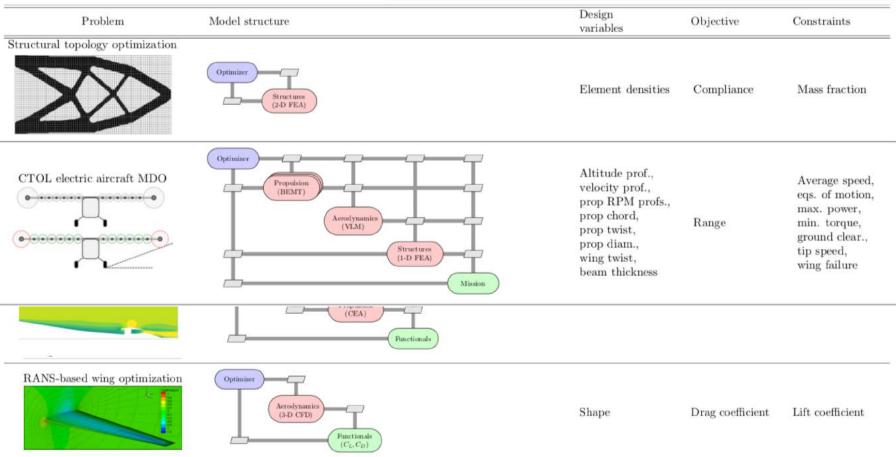
Discipline 2: $y_2(z_1, z_2, y_1) = \sqrt{y_1} + z_1 + z_2$

Multidisciplinary analysis (MDA) consists in solution of the following equations

$$R_1 = 0$$
 \rightarrow y_1 and y_2 solutions $R_2 = 0$



J. S. Gray, J. T. Hwang, J. R. A. Martins, K. T. Moore, and B. A. Naylor, "OpenMDAO: An Open-Source Framework for Multidisciplinary Design, Analysis, and Optimization," Structural and Multidisciplinary Optimization, 2019.

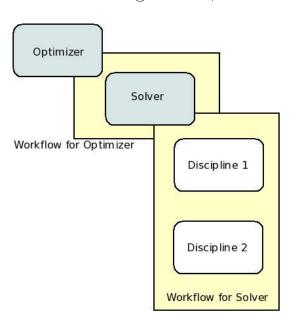


Multidisciplinary Feasible (MDF)

■ The MDF architecture is the most intuitive for engineers

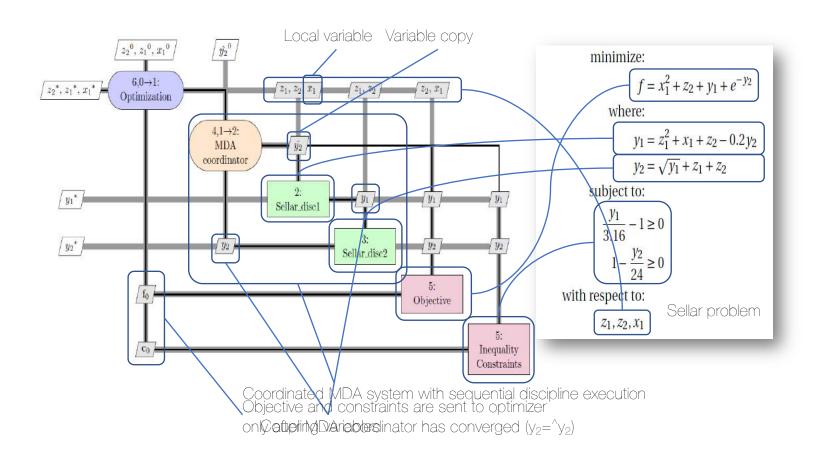
■ The optimization problem formulation is identical to the single discipline case, except the

disciplinary analysis is replaced by an MDA



MDF illustration on the Sellar problem:

MDF - Gauss-Seidel variant



Optimizer solver

- Optimizer

 Search direction

 Line search

 Converged?

 Analysis

 Gradient computation

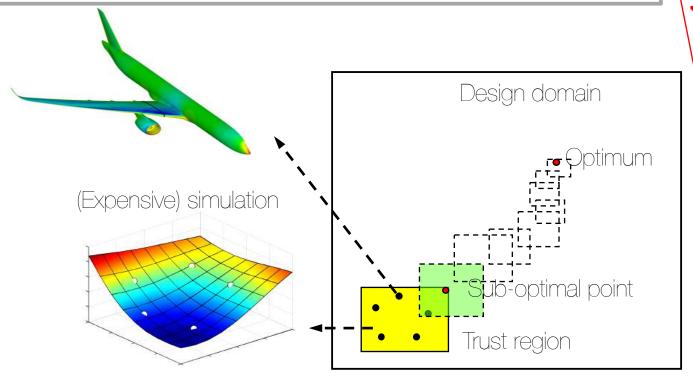
 Need to be accurate

- Derivative Fee Optimizer (DFO)
- Evolutionnary Strategies (ES)
- Surrogate based Optimizer (SBO) or Bayesian Optimization (BO)
-
- Gradient based Optimizer
- \rightarrow Computation of the derivatives of f(x) and $g_i(x)$ to iterate and satisfy the KKT optimality conditions
- → OpenMDAO focus on computation of sensitivities (adjoint vs direct)

$$\frac{\partial f}{\partial x_i}, \frac{\partial g}{\partial x_i}, \frac{\partial h}{\partial x_i}$$

SURROGATE MODELING (learning for Optimizing)

Jacobs, J. H., Etman, L. F. P., Van Keulen, F., & Rooda, J. E. (2004). Framework for sequential approximate optimization. Structural and Multidisciplinary Optimization, 27(5), 384-400.



Response surfaces, metamodels, surrogate models etc...

Outlines for today

1. MDO



3. ROM

A bit of History

Kriging (Pionneer)	Gaussian Processes (link with Al)
Developed by Daniel Krige – 1951; formalized by Georges Mathéron in the 60's (Mines Paris)	Neural network with infinite neurons tend to Gaussian Process 1994
Evaluation: minimize error variance	Evaluation: Marginal Likelihood

Krige, D. G., 1951, A statistical approach to some basic mine valuation problems on the Witwatersrand: J. Chem. Metal. Min. Soc. South Africa, v. 52, p. 119-139.

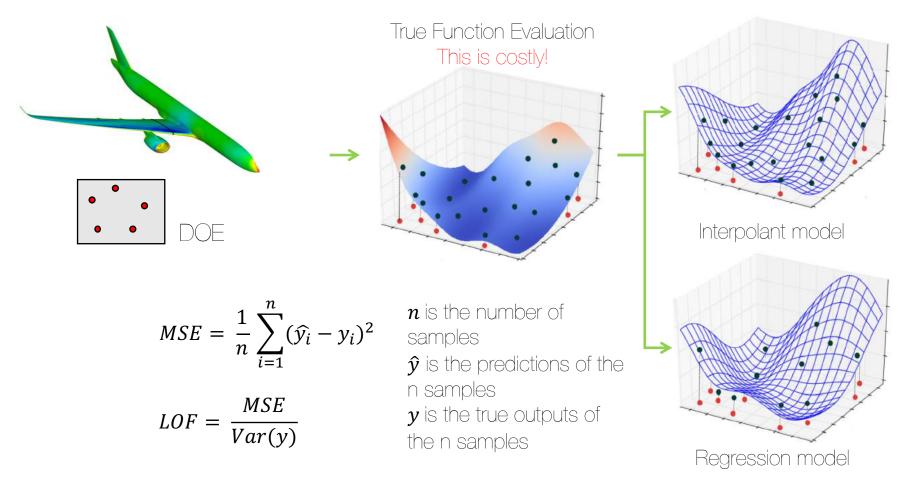
Matheron, G., 1963b, Principles of geostatistics: Economic Geol., v. 58, p. 1246-1266.



Neal, R. Priors for infinite networks. Tech. rep., University of Toronto, 1994. Williams, C. K. I., and Rasmussen, C. E. Gaussian processes for regression. Advances in Neural Information Processing Systems 8 (1996), 514–520.

http://extrapolated-art.com

Surrogate modeling Recipes



ML class 2010

A good starting point x₀=Rasmussen's book (ML)

A good starting point x_0 =Forrester's book (Aerospace)

https://drafts.distill.pub/gp/

A Practical Guide to Gaussian

Processes

Lengthscale

Signal (standard deviation)

Noise (standard deviation)

Click to add points

Gaussian processes are useful for probabilistic modeling of unknown functions. We characterize the behavior of the hyperparameters of Gaussian processes, which will guide us toward useful heuristics with respect to optimization and numerical stability.

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml

Gaussian Processes for Machine Learning

Engineering Design via Surrogate Modelling

A Practical Guide

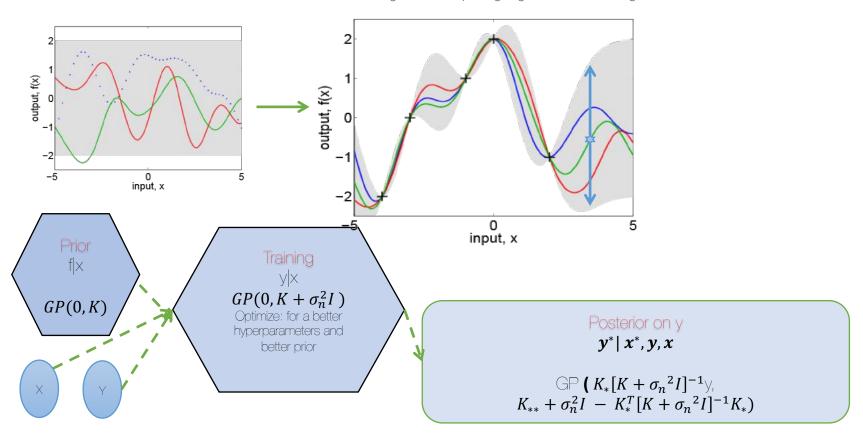
Alexander I. J. Forrester, András Sóbester and Andy J. Keane

University of Southampton, UK

ML class 2010

Gaussian Process (aka Kriging)

Image Source: http://mlg.eng.cam.ac.uk/teaching/4f13/1314/



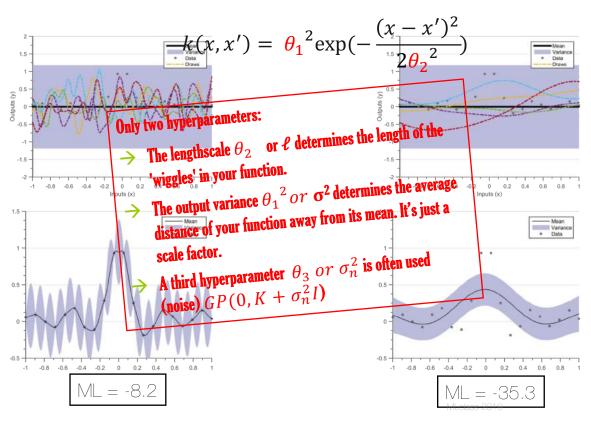
3/ Choose a Kernel/Construct Kxx Matrix view of Gaussian Process and Hyperparameters tuning 1/ Get your inputs/outputs data 2/ You wan to predict at x* $k(x, x') = \frac{\theta_1^2 \exp(-\frac{(x - x')^2}{2\theta_2^2})}{\|x\|_2^2}$ = $[K \times \times]$ $[K_{X^*X^*}]$ **-** | [Kx*x] and variance of estimate $m(x_*) = K_*[Kxx]^{-1}$ A compute mean $var(x_*, x_*') = K_{**} - K_*^T [Kxx]^{-1} K_*$

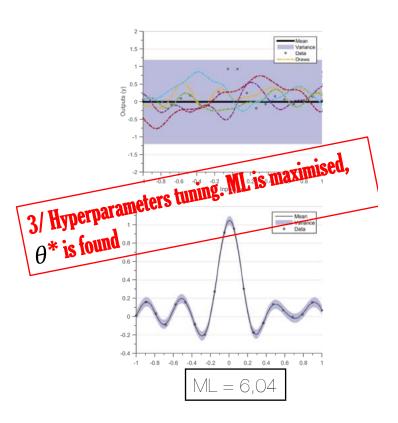
MI class 2019

Optimizing Marginal Likelihood (ML)

$$\mathsf{ML} = log(p(y|X,\theta)) = -\frac{1}{2}y^{\mathsf{T}}K^{-1}y - \frac{1}{2}log|K| - \frac{n}{2}log(2\pi)$$

It is a combination of data-fit term, a complexity penalty term and a normalization term



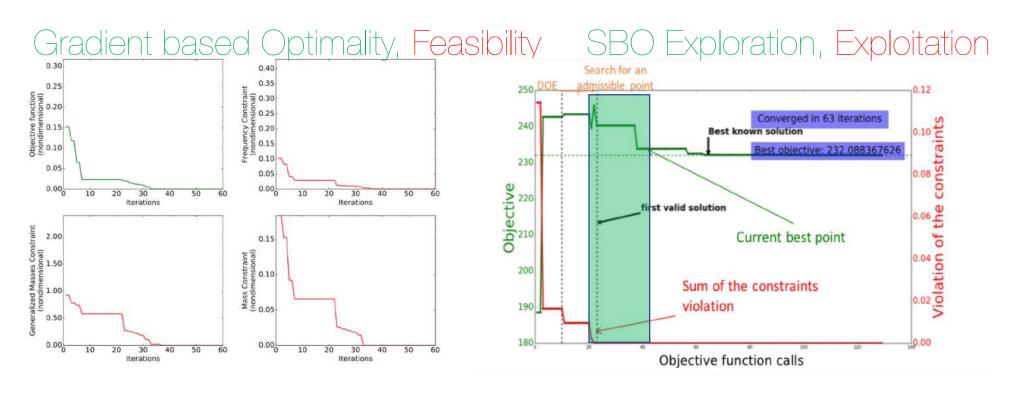


Need a demo?

https://distill.pub/2019/visual-exploration-gaussian-processes/



New paradigm for Surrogate Based Optimization (SBO)



Stopping criteria: tolfun, tolx, maxiter

Stopping criteria: Max Budget (Function calls)

ML class 2019

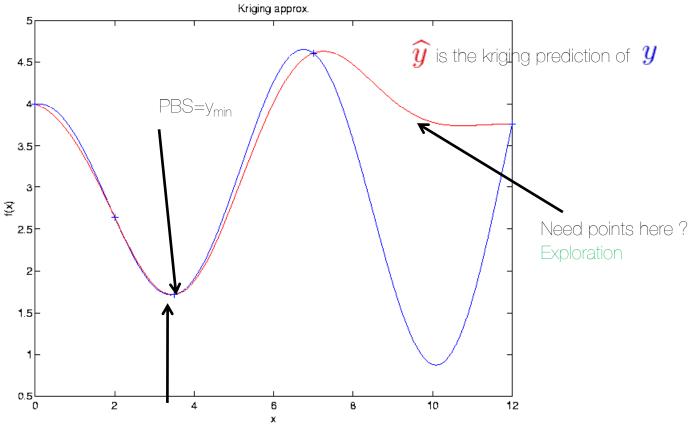
The goal is: find min of f(x) by sampling + and Kriging updating



We note the present best solution $(PBS=y_{min})$

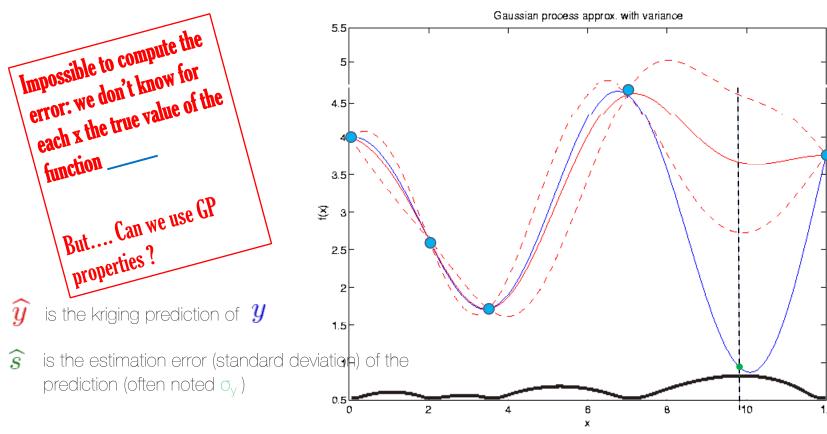
At every x there is some chance of improving on the PBS.

Then we ask: Assuming an improvement over the PBS, where is it likely be largest?



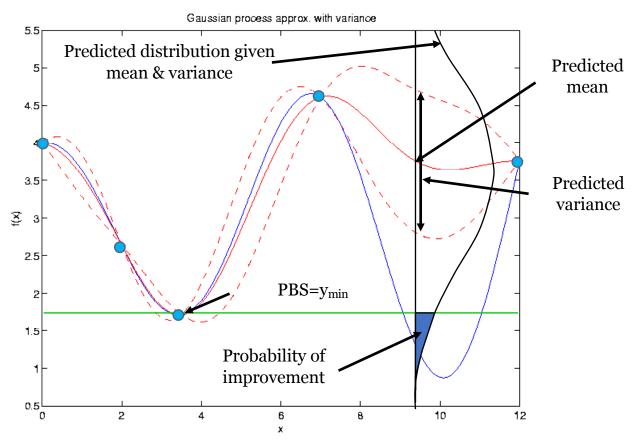
Exploitation may drive the optimization to a local optimum

In supervised mode ... have a look to max(RMSE)



PBS=y_{min}

Probability of improvement



Improvement ... explicitely

• Improvement :
$$I(\mathbf{x}) = \max \left(y_{min} - \hat{Y}(\mathbf{x}), 0 \right)$$

• Expected Improvement :

$$EI(x) = E[max (0, y_{min} - \hat{y}(x))]$$

$$E[I(\mathbf{x})] = \int_{-\infty}^{y_{min}} (y_{min} - \hat{y}) \varphi \left(\frac{y_{min} - \mu_{\hat{Y}}(\mathbf{x})}{\sigma_{\hat{Y}}(\mathbf{x})} \right) d\hat{y}$$

$$E[I(\mathbf{x})] = (y_{min} - \mu_{\hat{Y}}(\mathbf{x}))\Phi\left(\frac{y_{min} - \mu_{\hat{Y}}(\mathbf{x})}{\sigma_{\hat{Y}}(\mathbf{x})}\right) + \sigma_{\hat{Y}}(\mathbf{x})\varphi\left(\frac{y_{min} - \mu_{\hat{Y}}(\mathbf{x})}{\sigma_{\hat{Y}}(\mathbf{x})}\right)$$

Exploitation

Exploration

global optimum can be found because P[I(x)] = 0 when s = 0 so that there is no probability of improvement at a point which has already been sampled \rightarrow guarantees global convergence

 Φ : cumulative distribution function $\mathcal{N}(0,1)$ ϕ : probability density function

 $\mathcal{N}(0,1)$

*Jones, D. R., Schonlau, M., & Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. Journal of Global optimization, 13(4), 455-492.

Infill Criteria: max(Expected improvement)

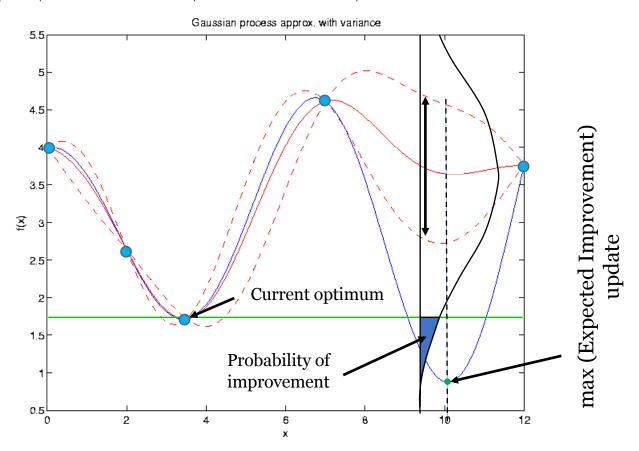
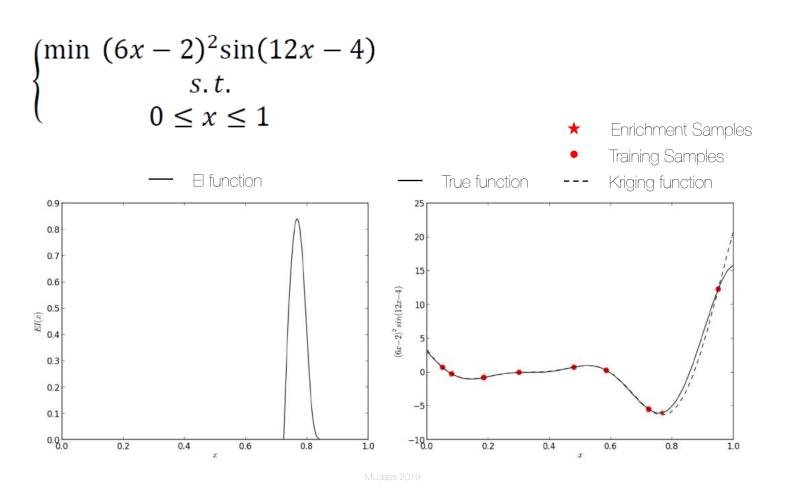


Illustration on 1D example



MOPTA



Problem formulation:

1 objective function to be minimized - mass

124 variables normalized to [0,1]

68 inequality constraints of form $g_i(x) \le 0$

Constraints well normalized: 0.05 means 5% over requirement, etc.

Test problem comes with the initial feasible point with objective ~251.07

A good performance would be comparable or better than derivative-free optimization algorithm - Powell's <u>COBYLA</u>:

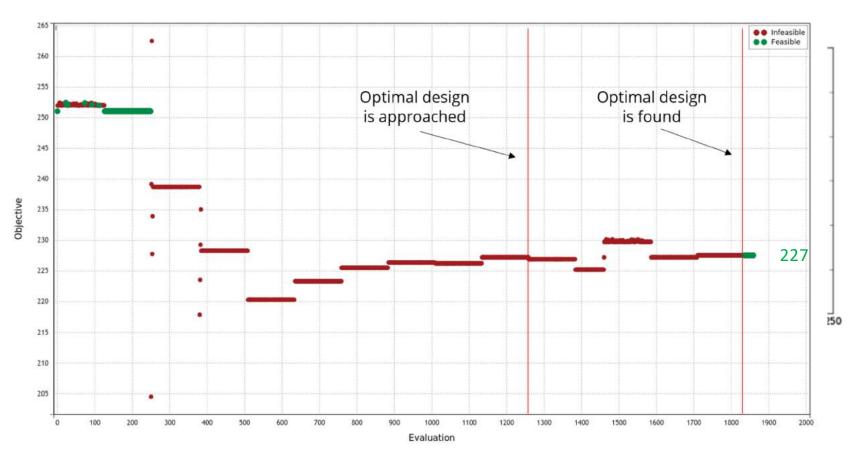
Number of evaluations = ~ 15 x Number of variables

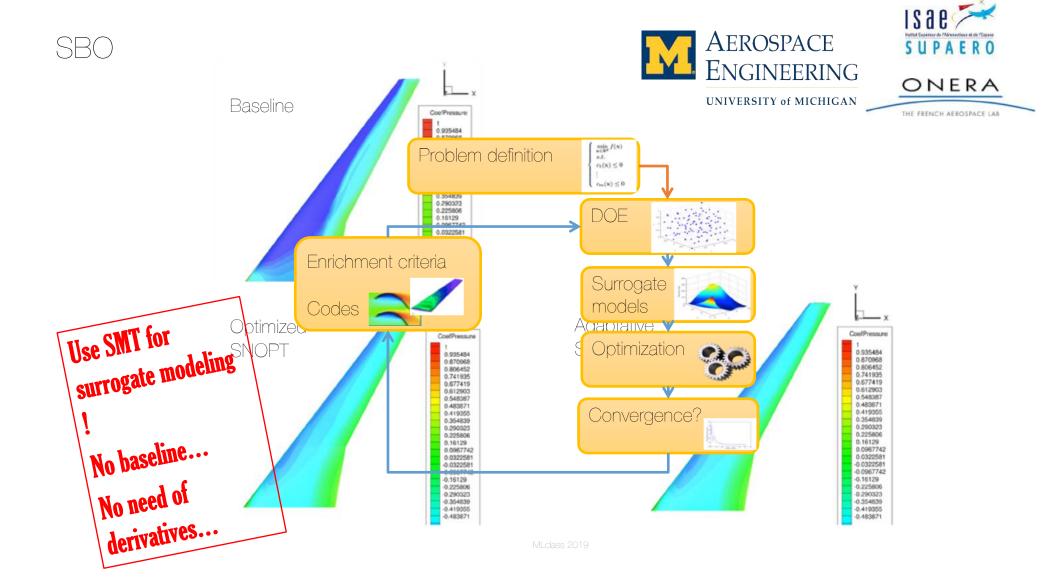
Fully feasible solution (no constraint violations)

Objective function value ≤ 228 (at least 80% of potential reduction)

"Anything better is exciting" - states Don Jones, the author of this benchmark.

MOPTA08 optimization history BO vs pSeven





Surrogate Model Toolbox: SMT



Table Of Contents

SMT: Surrogate Modeling Toolbox Focus on derivatives Documentation contents
• Indices and tables

Next topic

Getting started

This Page

Show Source

Quick search

SMT: Surrogate Modeling Toolbox

The surrogate model toolbox (SMT) is an open-source Python package consisting of libraries of surrogate modeling methods (e.g., radial basis functions, kriging), sampling methods, and benchmarking problems. SMT is designed to make it easy for developers to implement new surrogate models in a well-tested and well-document platform, and for users to have a library of surrogate modeling methods with which to use and compare methods.

The code is available open-source on GitHub.

Focus on derivatives

SMT is meant to be a general library for surrogate modeling (also known as metamodeling, interpolation, and regression), but its distinguishing characteristic is its focus on derivatives, e.g., to be used for gradient-based optimization. A surrogate model can be represented mathematically as

$$y = f(\mathbf{x}, \mathbf{xt}, \mathbf{yt}),$$

where $\mathbf{xt} \in \mathbb{R}^{ntx,nx}$ contains the training inputs, $\mathbf{yt} \in \mathbb{R}^{nt}$ contains the training outputs, $\mathbf{x} \in \mathbb{R}^{nx}$ contains the prediction inputs, and $y \in \mathbb{R}$ contains the prediction outputs. There are three types of derivatives of interest in SMT:

- 1. Derivatives (dy/dx): derivatives of predicted outputs with respect to the inputs at which the model is evaluated.
- Training derivatives (dytldxt): derivatives of training outputs, given as part of the training data set, e.g., for gradient-enhanced kriging.
- Output derivatives (dy/dyt): derivatives of predicted outputs with respect to training outputs, representing how the prediction changes if the training outputs change and the surrogate model is re-trained.

Not all surrogate modeling methods support or are required to support all three types of derivatives; all are optional.



https://github.com/SMTorg/SMT

Outlines for today

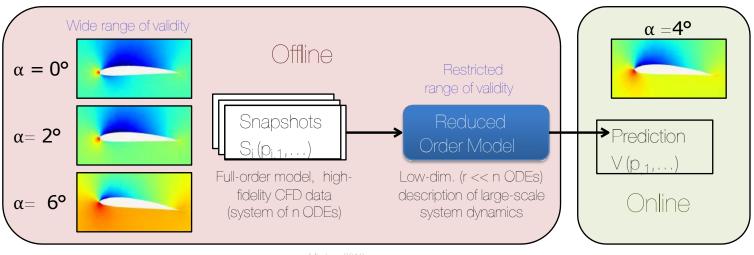
1. MDO 2. GP



How? Using POD (SVD, PCA, KLT)*

*
eceweb1.rutgers.edu/~orfanidi/ece525/svd.pdf

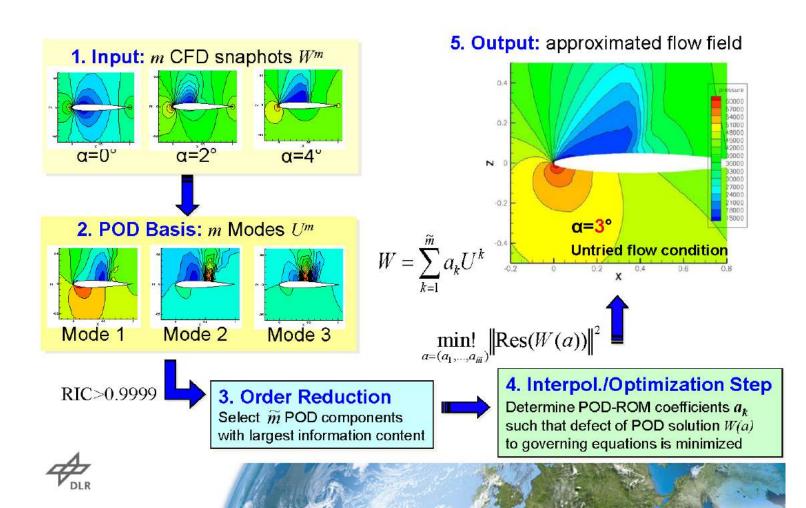
- Reduced order models operate on parameterized generated data (snapshots)
 - scalar quantities: lift, drag and moment coefficients CI, Cd, Cm
 - ullet surface quantities: pressure and shear stress, volume quantities: primitive variables $oldsymbol{
 ho}_{\!\scriptscriptstyle B} T$
- lacktriangle Parameters can be related to the flow (e.g. angle of attack lpha , Mach number \emph{M} or to the geometry



ML class 2019

https://www.aerogust.eu

POD-based Reduced Order Modeling

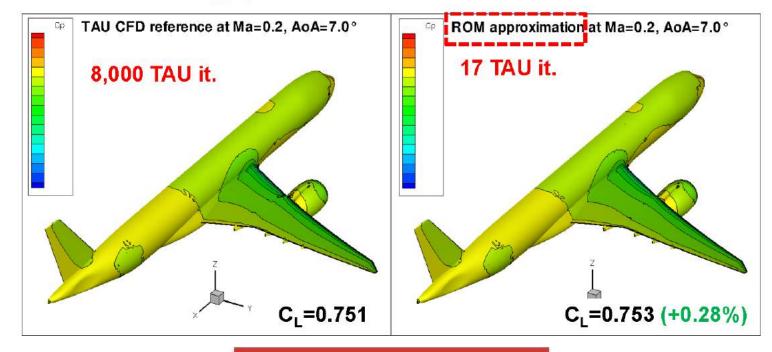


https://www.aerogust.eu

POD-based Reduced Order Modeling

- Industrial a/c config., grid size: 8898749
- 4 Snapshots at α = [-1°,0°,1°,2°] with TAU code
- Approximation at a = 7° (extrapolation)

Navier-Stokes ROM subsonic



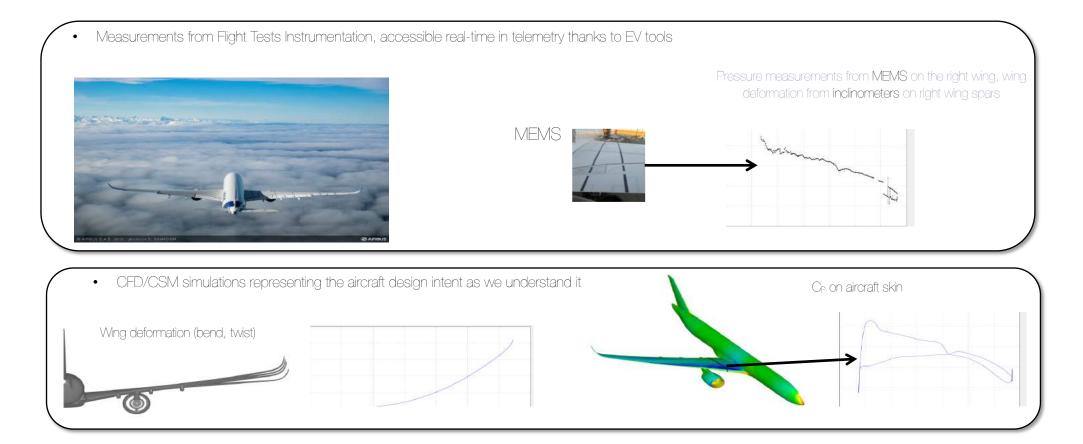


ROM speed-up by factor: 470

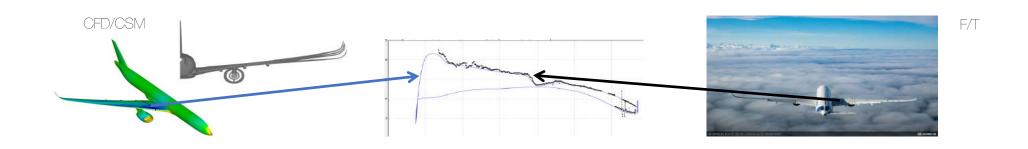
Example from a recent PhD thesis ...with your teacher Emmanuel Rachelson

https://github.com/ankitchiplunkar/thesis_isae

A350-1000: Interpolation of shock



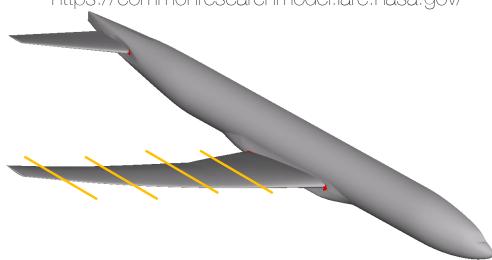
A350-1000 flight test analysis



- Comparing these two sources of data during flight tests enables us to check that the A/C behaves as expected, and that our understanding of its physics (aerodynamics, loads, wing shape) is correct.
- By doing this comparison live in telemetry, we can interact with the ongoing flight test, and optimize configurations (VC/DFS) if needed, for bringing the A/C behaviour as close as possible to design intent.
- These comparisons between CFD/CSM and Flight Test measurements have to be done at identical values of flight parameters: Mach, Alpha, Flight level, VC configuration... so we need to be able to plot instantly the CFD/CSM data for any given combination of these parameters.

Experimental dataset

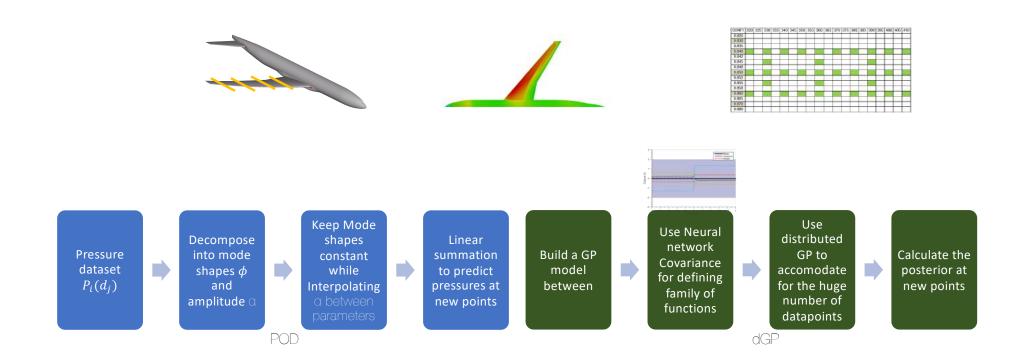
• https://commonresearchmodel.larc.nasa.gov/



- Simulations run using: Elsa kOmega-SST
- Same as the one proposed during drag prediction workshop
- · Gives better interaction between model fuselage and wing
- Alpha = [1 : 0.1 : 3] = **21 alphas**
- Mach = [0.84 : 0.005 : 0.86] = 5 machs
- yLocationCuts = [6.03, 11.99, 17.76, 27.85]
- Wanted to be close to (as used in the drag prediction workshop)
 - y/b = [0,105, 0,37, 0,5024, 0,8456]
 - Building a model between alpha, mach and x

Chiplunkar, Ankit and Bosco, Elisa and Morlier, Joseph Gaussian Process for Aerodynamic Pressures Prediction in Fast Fluid Structure Interaction Simulations. (2017) In: 12th World Congress on Structural and Multidisciplinary Optimization, 5 June 2017 - 7 June 2017 (Braunschweig, Germany).

Different models

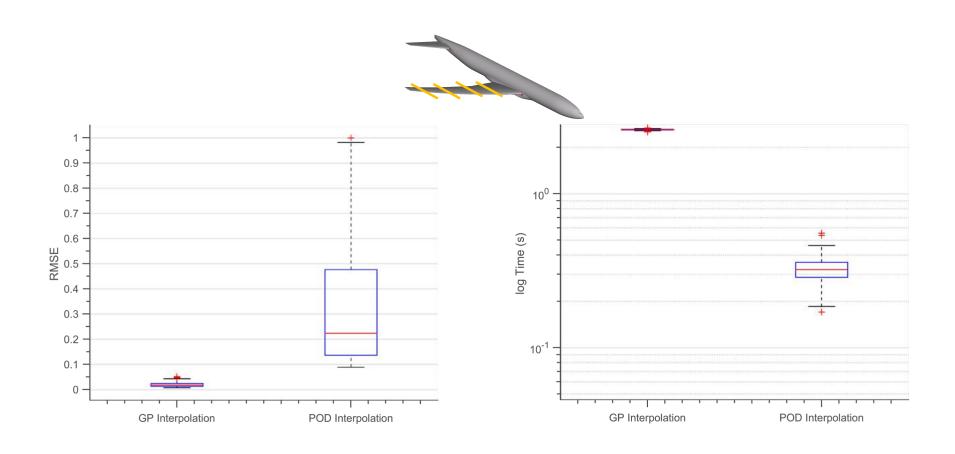


References:

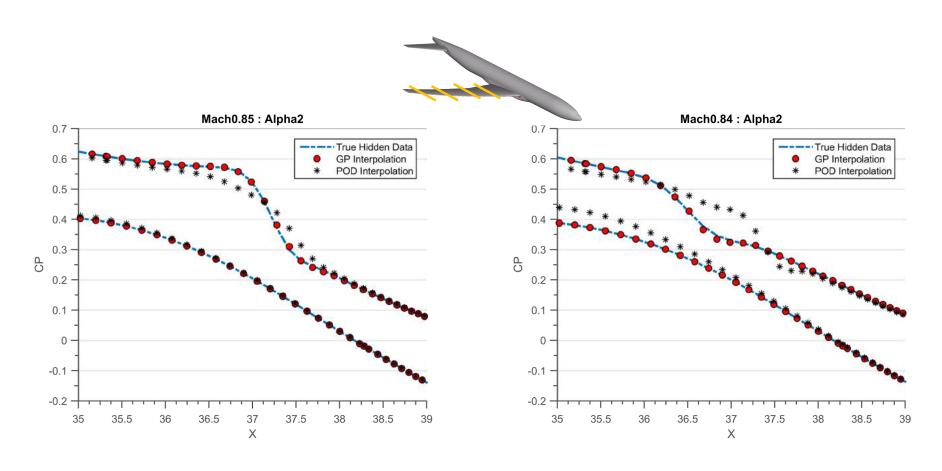
Proper orthogonal decomposition extensions and their applications in steady aerodynamics. Master of Engineering in High Performance Computation for Engineered Systems (HPCES), 2003.

Radford M Neal. Bayesian learning for neural networks, volume 118. Springer Science & Business Media, 2012

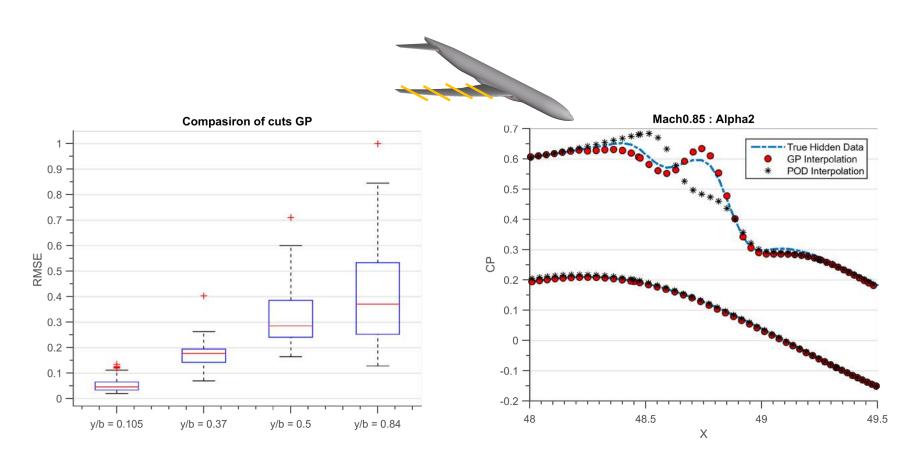
Comparison of results: Cut1 (y/B = 0,105)



Comparison of results: Cut1 (y/B = 0,105)

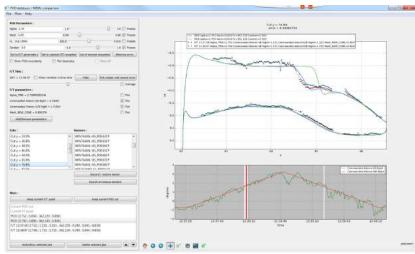


Comparison across Cuts



A350-1000: CP interpolation in telemetry room





- 1. Once validated the fast CFD/CSM model can be used in the telemetry room
- 2. Generally this task used to take 3 weeks now it is done in 3 seconds in the telemetry room
- 3. Opened many other use cases in the telemetry room, assisting performance optimization, verifying values of flight parameters etc
- 4. PODCraft selected for EG Innovation Recognition Event 2017

CONFIDENTIAL

Conclusion

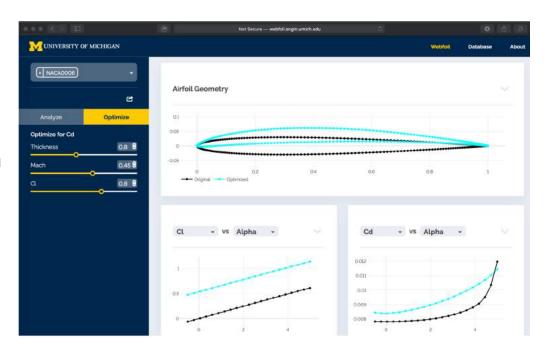
• Machine Learning technics are widely used in MDO to speed up the process (Offline/Online)

• A nice example is the XFOIL « regression » with SMT

http://webfoil.engin.umich.edu

• Go deeper with BO before the Python Practice (23/10)

http://krasserm.github.io/2018/03/21/bayesianoptimization/



Webfoil is a free, web-based airfoil design tool

EGO on SMT

Recent Papers on this topic



Bouhlel, M. A., Bartoli, N., Otsmane, A., & Morlier, J. (2016). Improving kriging surrogates of high-dimensional design models by Partial Least Squares dimension reduction. Structural and Multidisciplinary Optimization, 53(5), 935-952.

Bouhlel, M. A., Bartoli, N., Otsmane, A., & Morlier, J. (2016). An improved approach for estimating the hyperparameters of the kriging model for high-dimensional problems through the partial least squares method. Mathematical Problems in Engineering, 2016.



Bouhlel, M., Bartoli, N., Regis, R. G., Otsmane, A., & Morlier, J. (2018). Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method. Engineering Optimization, 1-16.

Bouhlel, M. A., Hwang, J. T., Bartoli, N., Lafage, R., Morlier, J., & Martins, J. R. (2019). A Python surrogate modeling framework with derivatives. Advances in Engineering Software, 102662.



Bartoli, N., Lefebvre, T., Dubreuil, S., Olivanti, R., Priem, R., Bons, N., ... & Morlier, J. (2019). Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design. Aerospace Science and technology, 90, 85-102.

Chiplunkar A., Rachelson E., Colombo M., Morlier J. (2017) Approximate Inference in Related Multi-output Gaussian Process Regression. In: Fred A., De Marsico M., Sanniti di Baja G. (eds) Pattern Recognition Applications and Methods. ICPRAM 2016. Lecture Notes in Computer Science, vol 10163. Springer, Cham



Lecture Notes in Computer Science