SIAM CSE21, 1-5 MARCH 2021

SISSA, INTERNATIONAL SCHOOL FOR ADVANCED STUDIES, TRIESTE, ITALY

MARCO TEZZELE, NICOLA DEMO, GIANLUIGI ROZZA

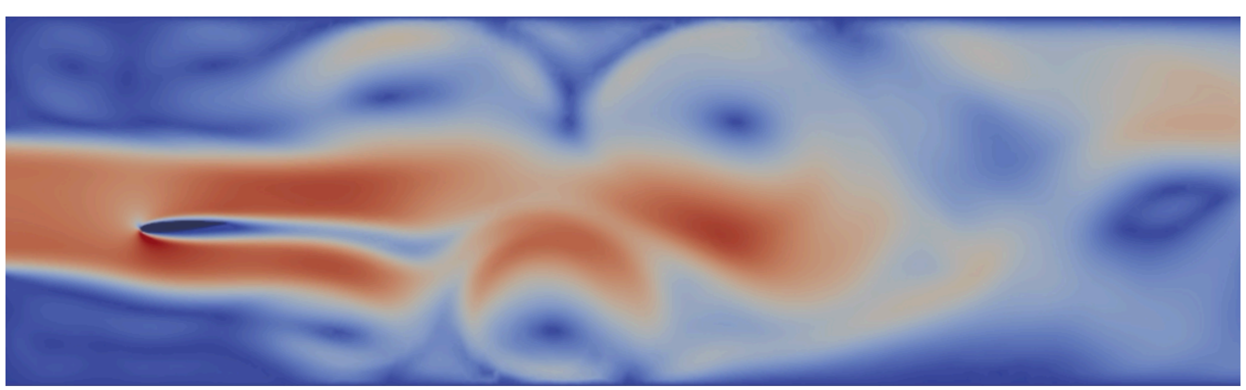
ENHANCING MODEL ORDER REDUCTION BY MACHINE LEARNING

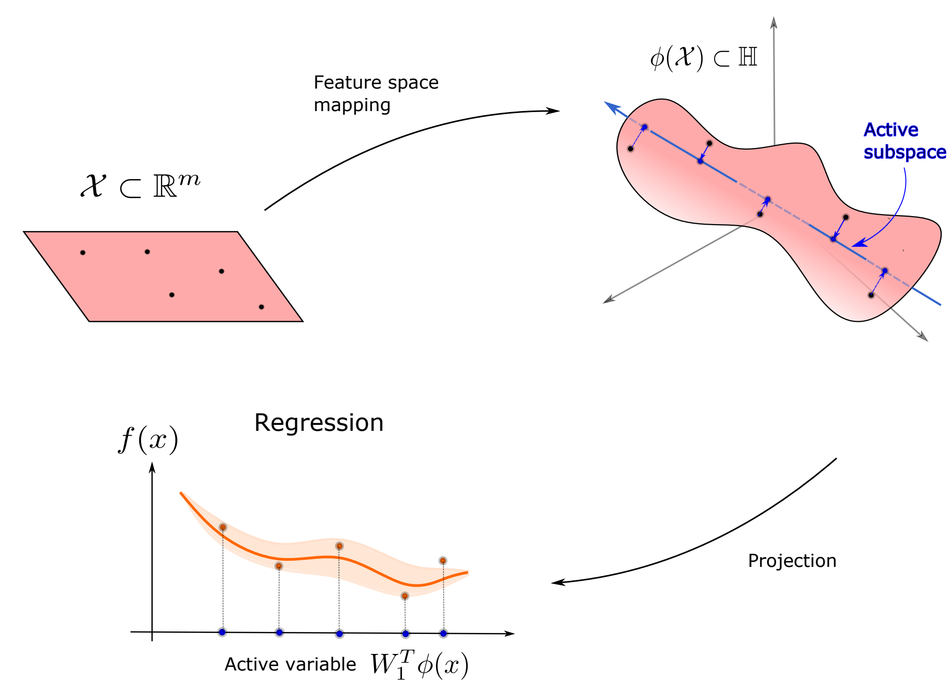
Outline

In this presentation we are going to show a nonlinear extension of the classical linear Active Subspaces (AS) methodology, based on reproducing kernel Hilbert spaces. We apply this new method to some benchmarks and to a CFD application using Discontinuous Galerkin method. Finally we show a shape optimization problem solved exploiting AS.

In particular, we are going to present:

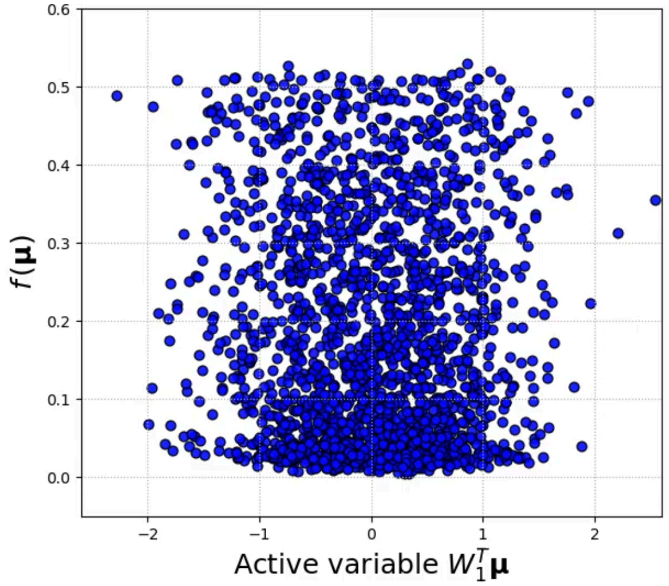
- An introduction to linear **Active Subspaces (AS)** for reduction in parameter space.
- A new **Kernel-based Active Subspaces (KAS)** extension, with applications.
- A coupling between **Active Subspaces and Genetic Algorithm (ASGA)** to enhance mono-objective optimization performance.

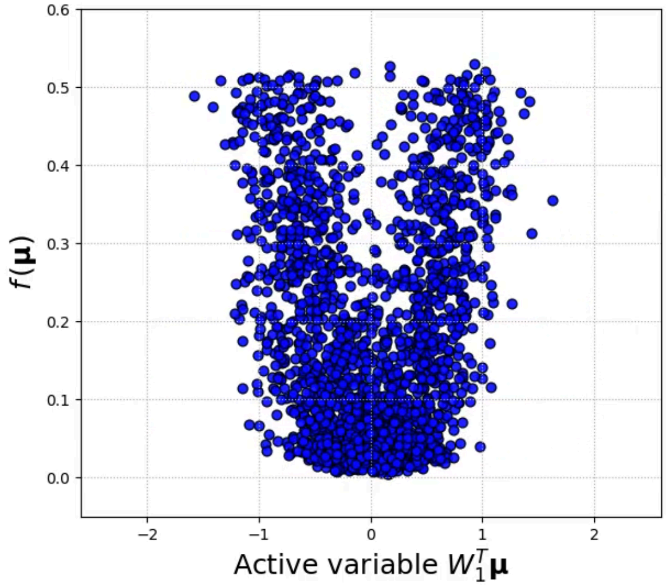


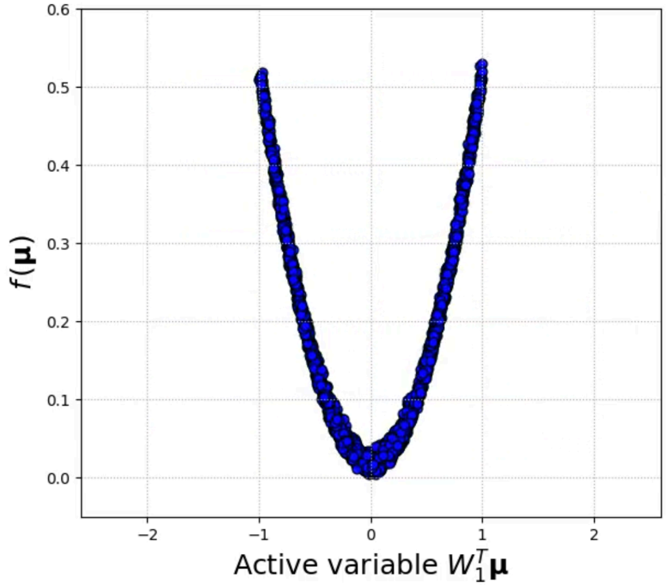


#ParameterSpaceReduction

From Linear Active Subspaces to a Kernel-based Approach







Active Subspaces formulation

Assumption: in many cases the dimension of the parametrized problem is **only artificially high**.

- **Active subspaces (AS)** property identifies a set of important directions in the space of all inputs [Constantine, 2015].

- **H** is the uncentered **covariance matrix** of the gradients of **f**. It is symmetric and positive semidefinite.

$$H = \int_{\mathcal{X}} (D_{\mathbf{x}} f(\mathbf{x}))^T R_V(\rho) (D_{\mathbf{x}} f(\mathbf{x})) \, d\rho(\mathbf{x})$$

- We are looking for a spectral gap in order to identify the **dimension r** of the active subspace. The **AS is the range of the first r eigenvectors**.

$$\tilde{H}\mathbf{v}_i = \lambda_i \mathbf{v}_i \qquad \forall i \in \{1, \dots, m\},$$
$$W_1 = (\mathbf{v}_1, \dots, \mathbf{v}_r), \quad W_2 = (\mathbf{v}_{r+1}, \dots, \mathbf{v}_m)$$

- With the basis identified we can map the parameters from the original full parameter space onto the **active and inactive subspace**, respectively:

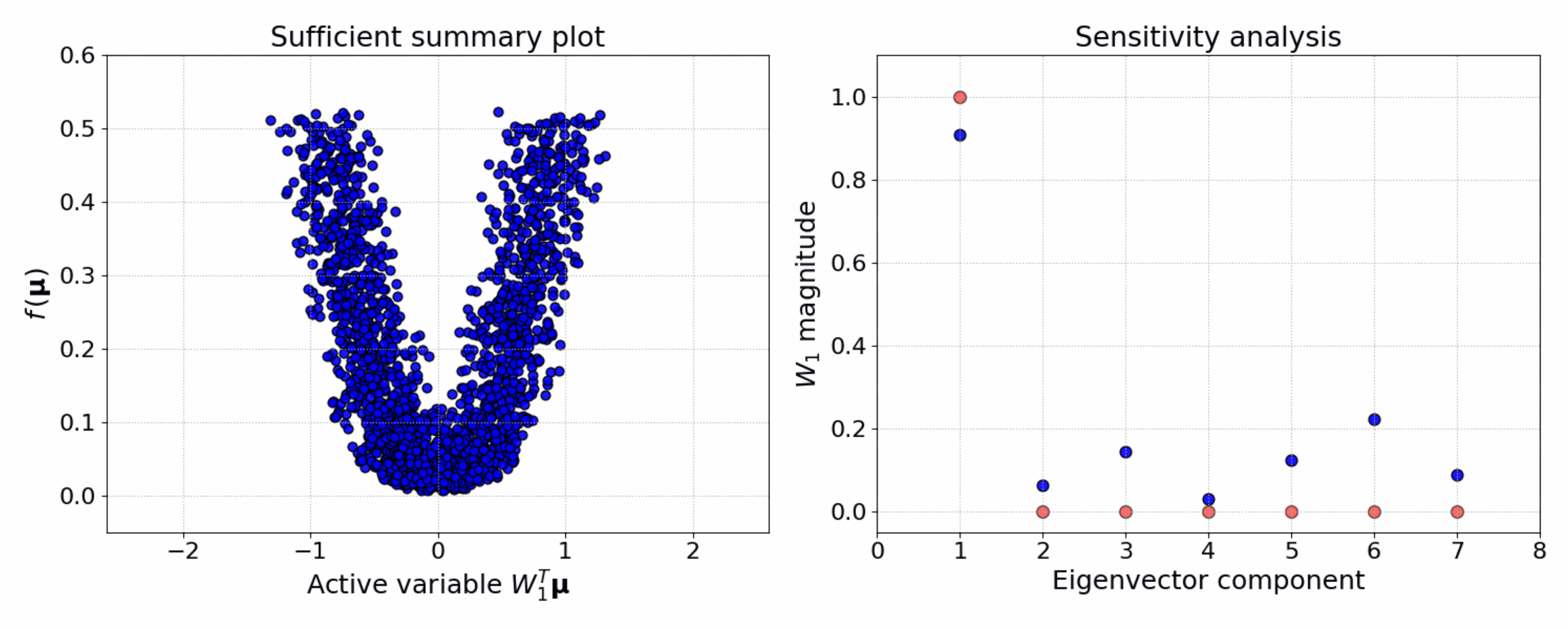
$$\mathbf{y} = W_1^T \mathbf{x} \in \mathbb{R}^r \qquad \mathbf{z} = W_2^T \mathbf{x} \in \mathbb{R}^{m-r}$$

If the simulation's prediction does not change as the inputs move along a particular direction, then we can safely ignore that direction in the parameter study.

Active Subspaces: linear parameter space reduction

- With a proper rotation of the input parameters domain, based on spectral properties, we are able to unveil lower dimensional behaviour of the function of interest.

- Reduction of the number of parameters through **active subspaces** has been proven useful in the case of small high fidelity solution databases in [1].

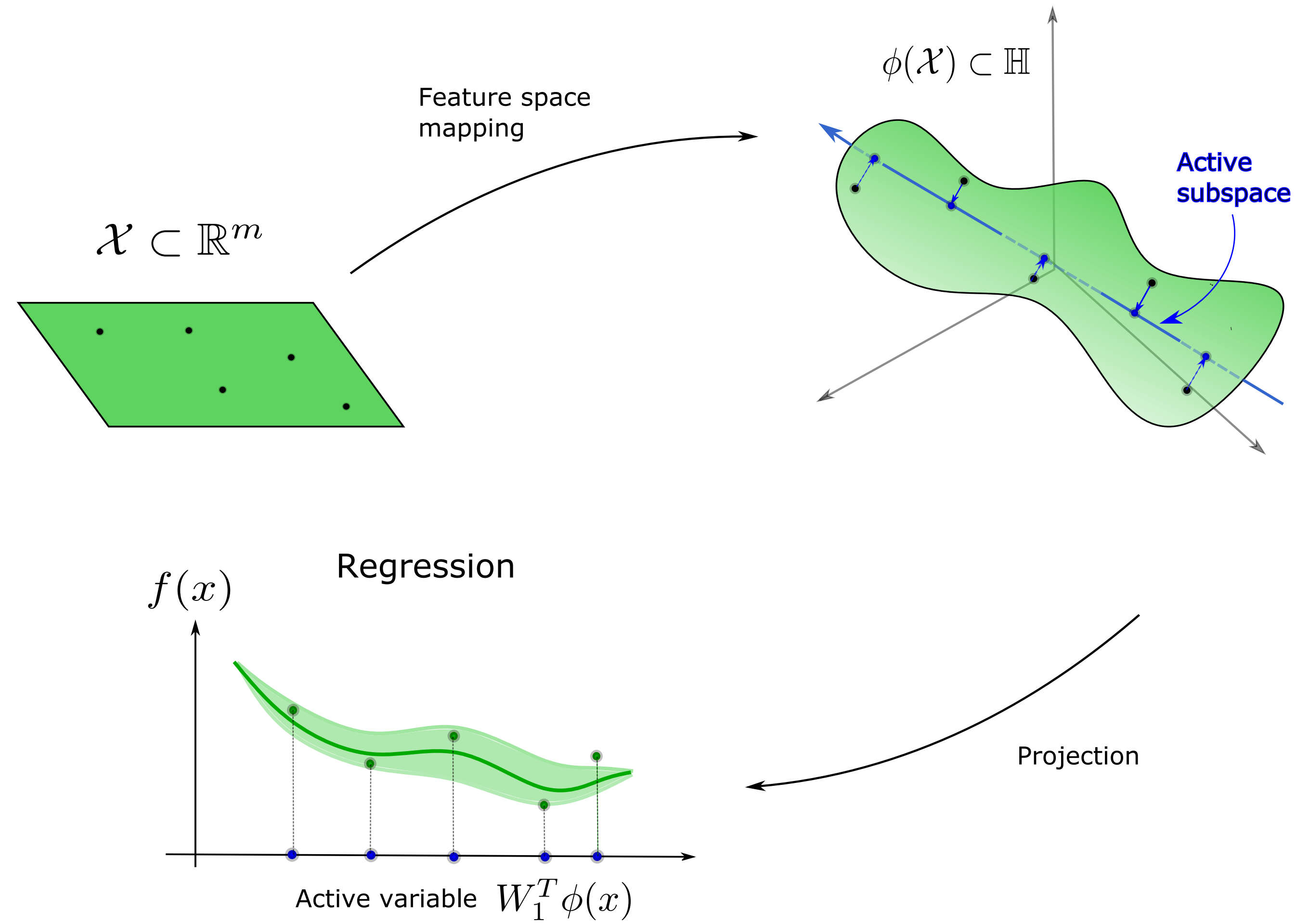


[1] N. Demo, M. Tezzele, and G. Rozza. "*A non-intrusive approach for proper orthogonal decomposition modal coefficients reconstruction through active subspaces*". Comptes Rendus Mécanique de l'Académie des Sciences, DataBEST 2019 Special Issue 347, 873–881, 2019.

Kernel-based Active Subspaces extension [2]

The main idea is to map the input parameters using an immersion from the parameter space to an infinite-dimensional Hilbert space:

$$\begin{tikzcd} \mathcal{X} \subset \mathbb{R}^m \arrow[r, "\phi"] \arrow[dr, "f"'] & \phi(\mathcal{X}) \subset \mathbb{H} \arrow[d, "\tilde{f}"] \\ & V \end{tikzcd}$$



The covariance matrix becomes:

$$\begin{aligned} \tilde{\mathcal{H}} &= \int_{\phi(\mathcal{X})} \left[(D_{\mathbf{z}}\tilde{f})^T(\mathbf{z}) \right] R_V \left[(D_{\mathbf{z}}\tilde{f})(\mathbf{z}) \right] d\mu(\mathbf{z}) \\ &= \int_{\mathcal{X}} \left[(D_{\mathbf{z}}\tilde{f})^T(\phi(\mathbf{x})) \right] R_V \left[(D_{\mathbf{z}}\tilde{f})(\phi(\mathbf{x})) \right] d\mathcal{L}_{\mathbf{X}}(\mathbf{x}) \\ &\approx \frac{1}{M}\sum_{i=1}^{M} \left[(D_{\mathbf{z}}\tilde{f})^T(\phi(\mathbf{x}_i)) \right] R_V \left[(D_{\mathbf{z}}\tilde{f})(\phi(\mathbf{x}_i)) \right] \end{aligned}$$

The choice for the feature map is linked to the theory of Reproducing Kernel Hilbert Spaces (RKHS) and it is defined as

$$\mathbf{z} = \phi(\mathbf{x}) = \sqrt{\frac{2}{D}}\,\sigma_f \, \cos(W\mathbf{x} + \mathbf{b})$$

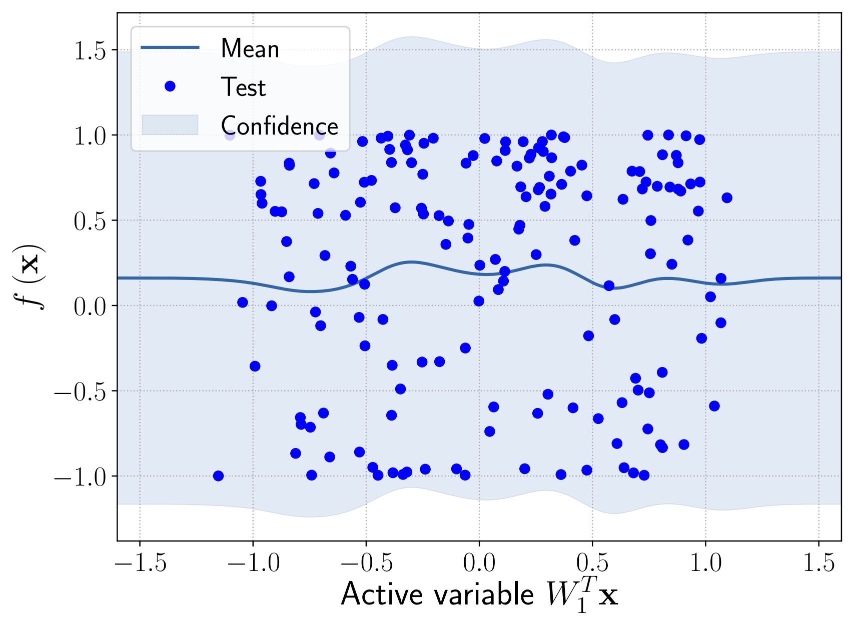
[2] F. Romor, M. Tezzele, A. Lario, and G. Rozza. "*Kernel-based Active Subspaces with application to CFD problems using Discontinuous Galerkin method*". Submitted, 2020. arXiv:2008.12083

Radial symmetric functions case

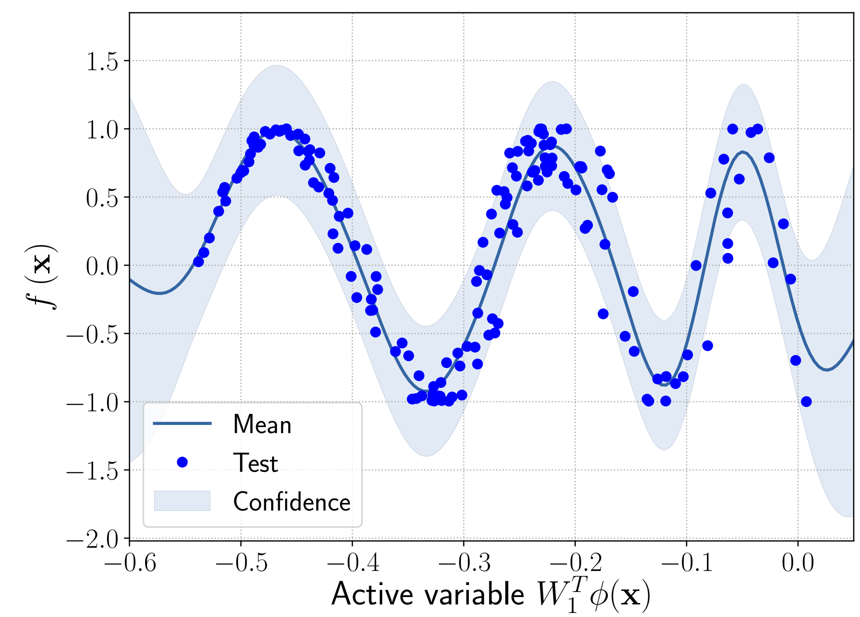
Comparison between the sufficiency summary plots obtained from the application of AS and KAS methods for

$$f : [-3, 3]^2 \subset \mathbb{R}^2 \to \mathbb{R},$$
$$f(\mathbf{x}) = \sin(\|\mathbf{x}\|^2).$$

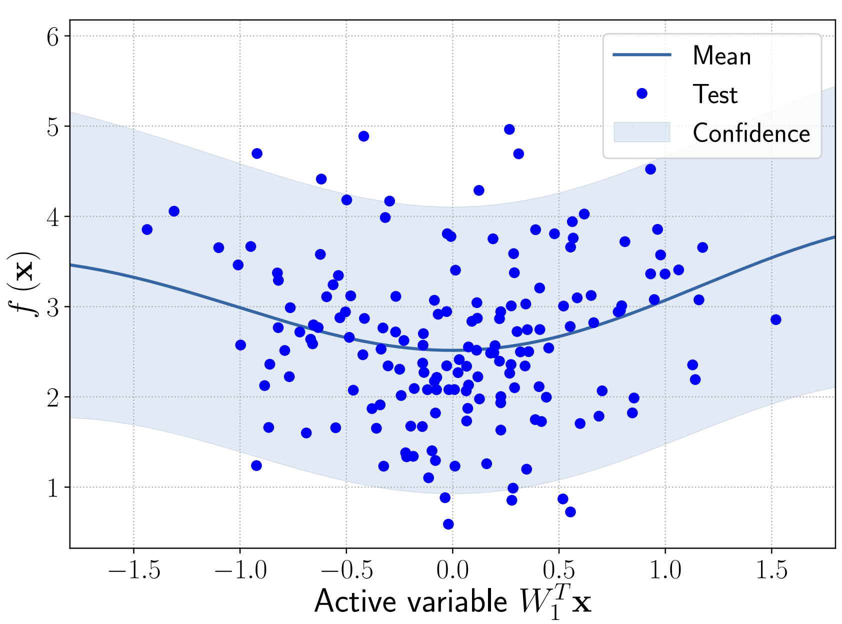
AS



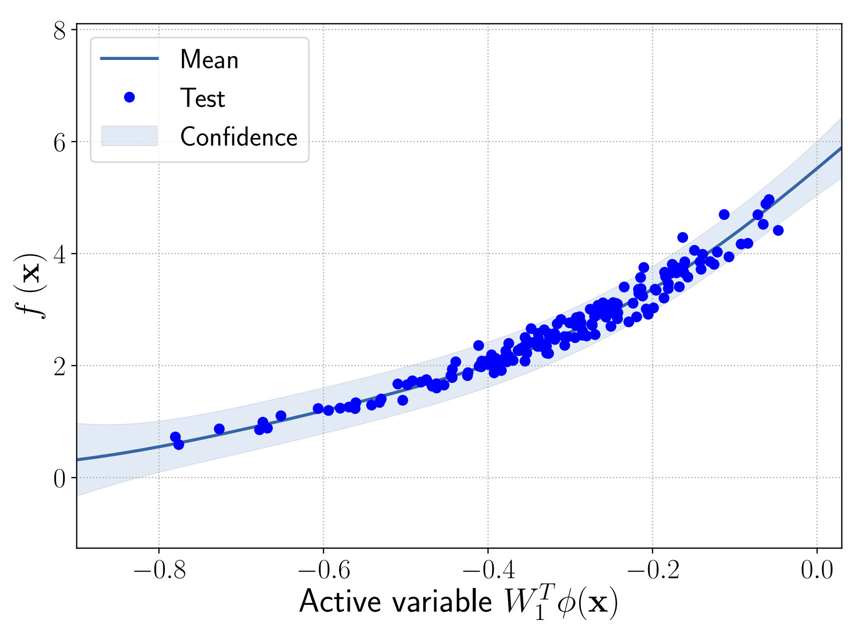
KAS



AS



KAS



Comparison between the sufficiency summary plots obtained from the application of AS and KAS methods for

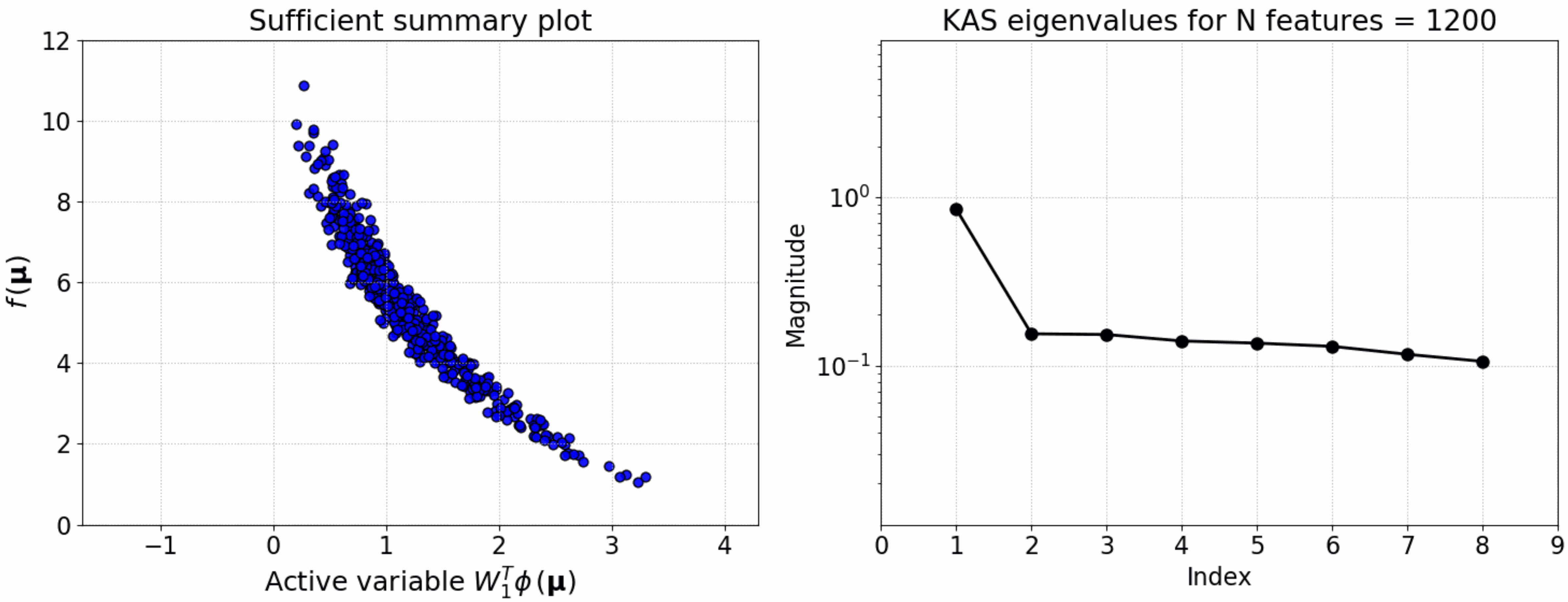
$$f : [-1, 1]^8 \subset \mathbb{R}^8 \to \mathbb{R},$$
$$f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$$

Sensitivity over the number of features

We can see how the dimension of the features space affects the spectral decay and the projection along the corresponding one-dimensional active variable.

We test it over the following hyperparaboloid model function:

$$f : [-2, 2]^8 \subset \mathbb{R}^8 \to \mathbb{R}, \qquad f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$$



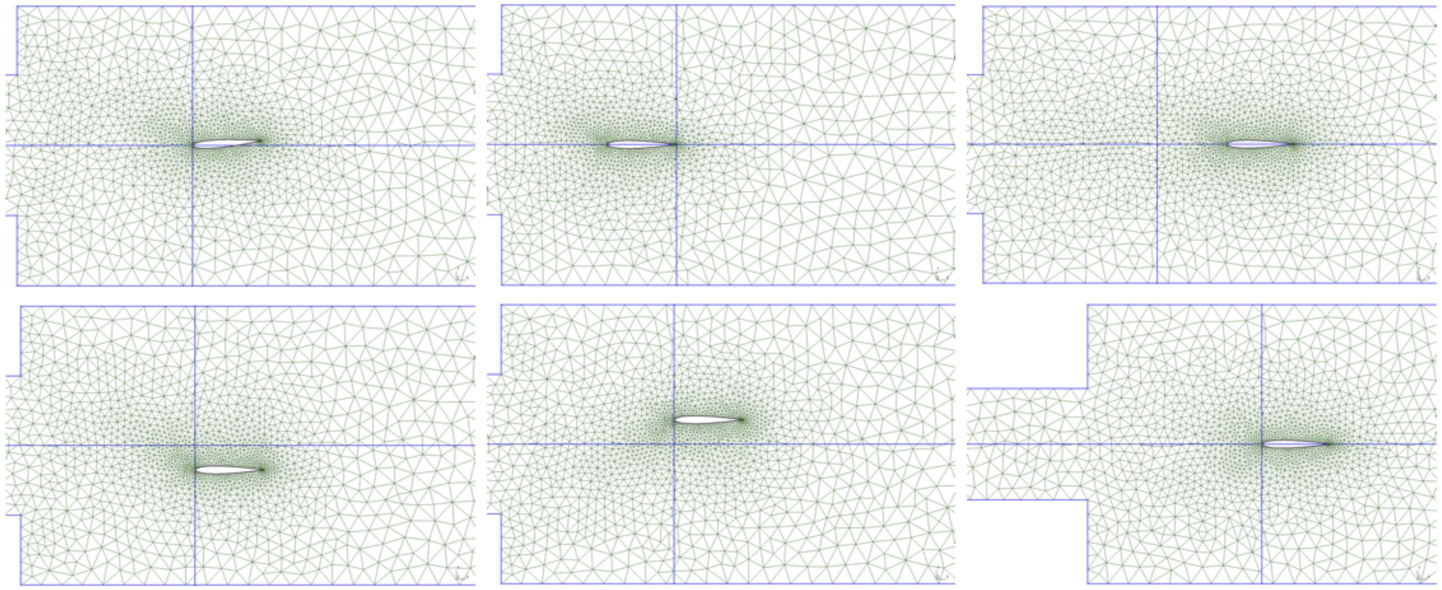
A CFD application using the DG method

The aerodynamic quantities we are interested in are the lift and drag coefficients in the incompressible case computed from

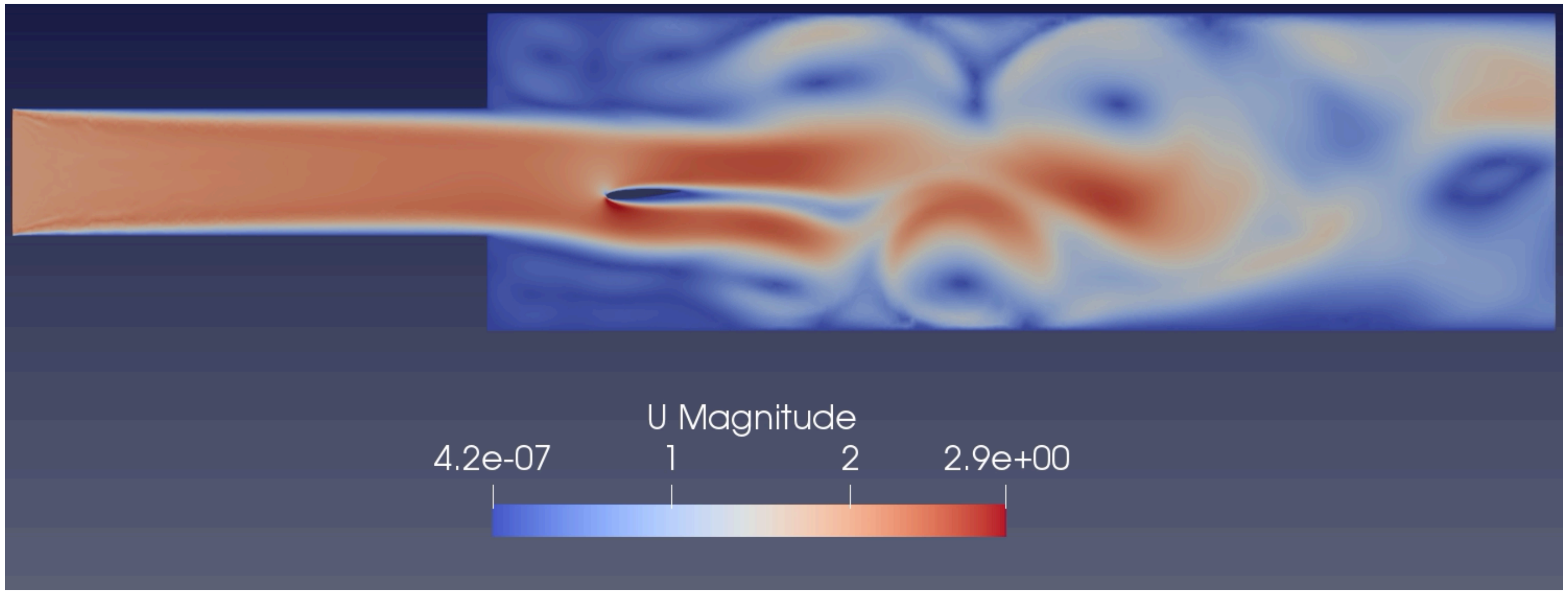
$$f = \oint_{\Gamma} p\mathbf{n} - \nu \left(\nabla \mathbf{u} + \nabla \mathbf{u}^T \right) \mathbf{n} \, ds.$$

$$C_D = \frac{f \cdot \mathbf{e}_1}{\frac{1}{2}|\mathbf{u}_0|^2 A_{\text{ref}}}, \qquad C_L = \frac{f \cdot \mathbf{e}_2}{\frac{1}{2}|\mathbf{u}_0|^2 A_{\text{ref}}}$$

Some corner cases of the parameter space



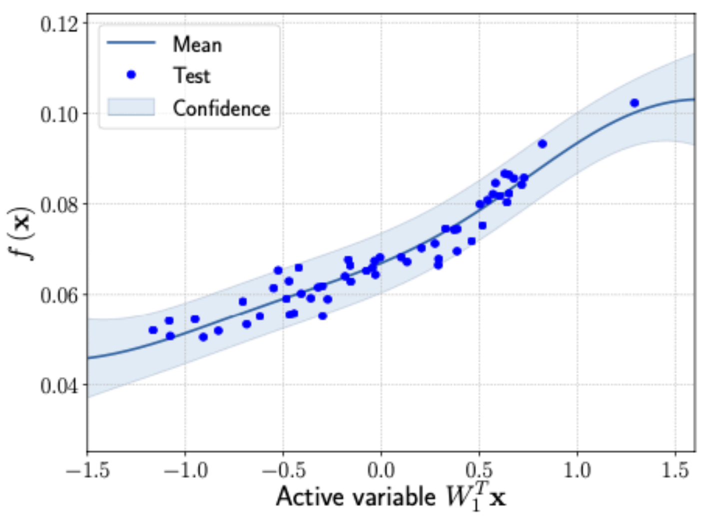
	ν	U	x_0	y_0	α	y^+	y^-
Lower bound	0.00036	0.5	-0.099	-0.035	0	-0.02	-0.02
Upper bound	0.00060	2	0.099	0.035	0.0698	0.02	0.02

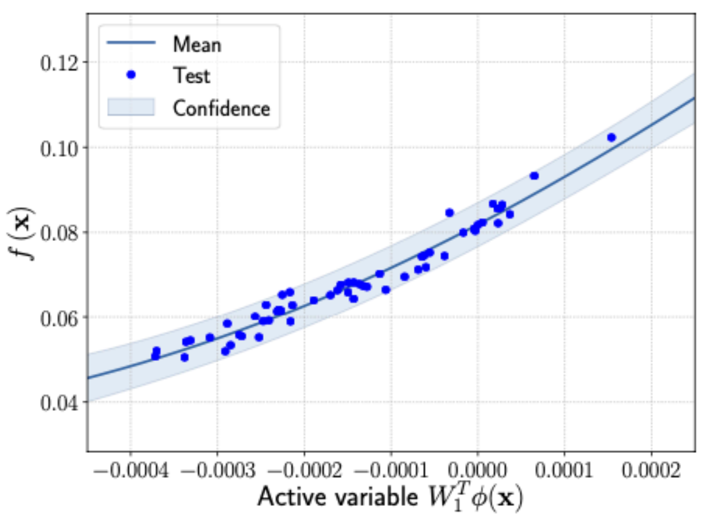


A CFD application using the DG method

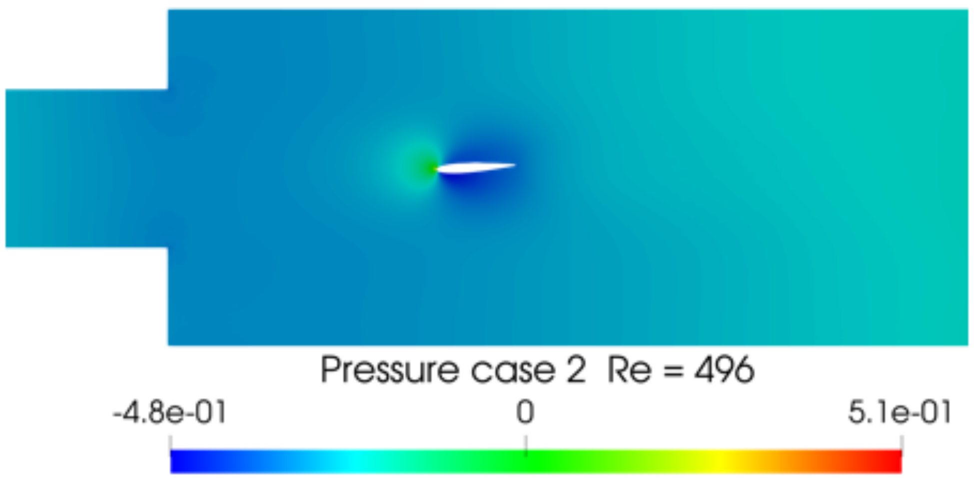
The relative gain using KAS for the drag coefficient is 19.2% on average when using the Beta spectral measure for the definition of the feature map. For the lift coefficient the gain is 7% on average. This result could be due to the higher noise in the evaluation of the CL. With a 2-dimensional Gaussian process response surface the gain reaches 14.6%.

Sufficient summary plot for AS and KAS for the lift coefficient





Pressure field



Method	Dim	Feature space dim	Lift spectral distribution	RRMSE Lift	Drag spectral distribution	RRMSE Drag
AS	1	-	-	0.37 ± 0.09	-	0.268 ± 0.032
KAS	1	1500	$\mathcal{N}(0, \lambda I_d)$	$\mathbf{0.344} \pm 0.048$	Beta(α, β)	$\mathbf{0.218} \pm 0.045$
AS	2	-	-	0.384 ± 0.073	-	0.183 ± 0.027
KAS	2	1500	$\mathcal{N}(0, \lambda I_d)$	$\mathbf{0.328} \pm 0.071$	Beta(α, β)	$\mathbf{0.17} \pm 0.02$

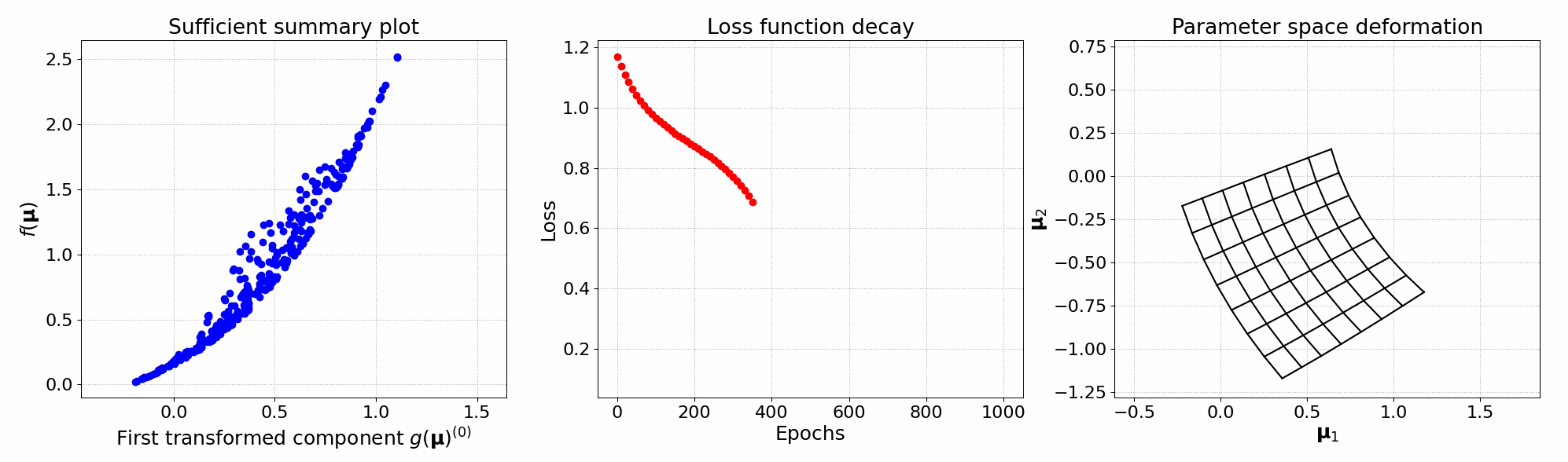
F. Romor, M. Tezzele, A. Lario, and G. Rozza. "*Kernel-based Active Subspaces with application to CFD problems using Discontinuous Galerkin method*". Submitted, 2020. arXiv:2008.12083

Open source software for parameter space reduction

ATHENA: Advanced Techniques for High dimensional parameter spaces to Enhance Numerical Analysis. Available at github.com/mathLab/ATHENA.

This Python package implements several parameter space dimensionality reduction techniques which can be coupled with other ROMs. Such methods include *Active Subspaces*, *Kernel-based Active Subspaces*, and *Nonlinear Level Set Learning* [*].

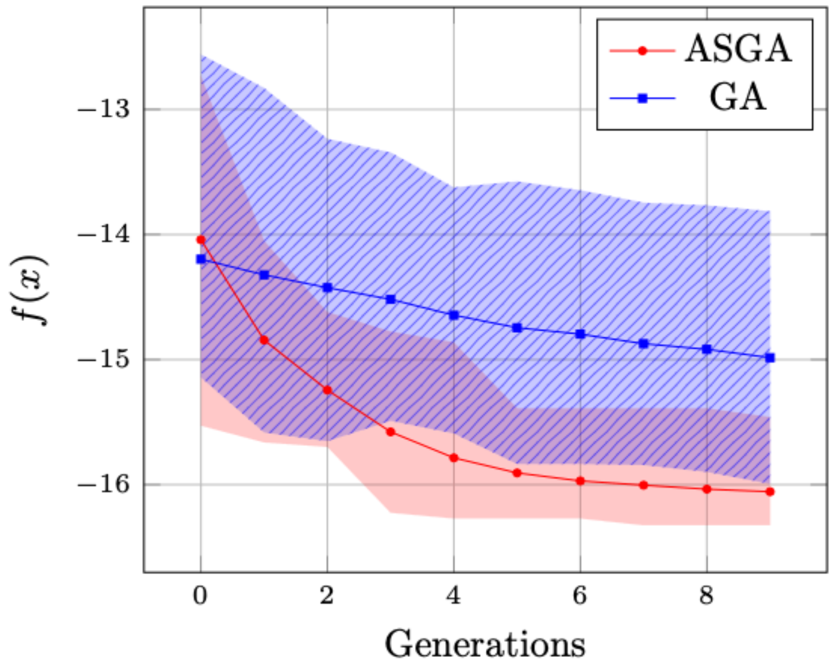
It is developed and maintained by Francesco Romor and myself.

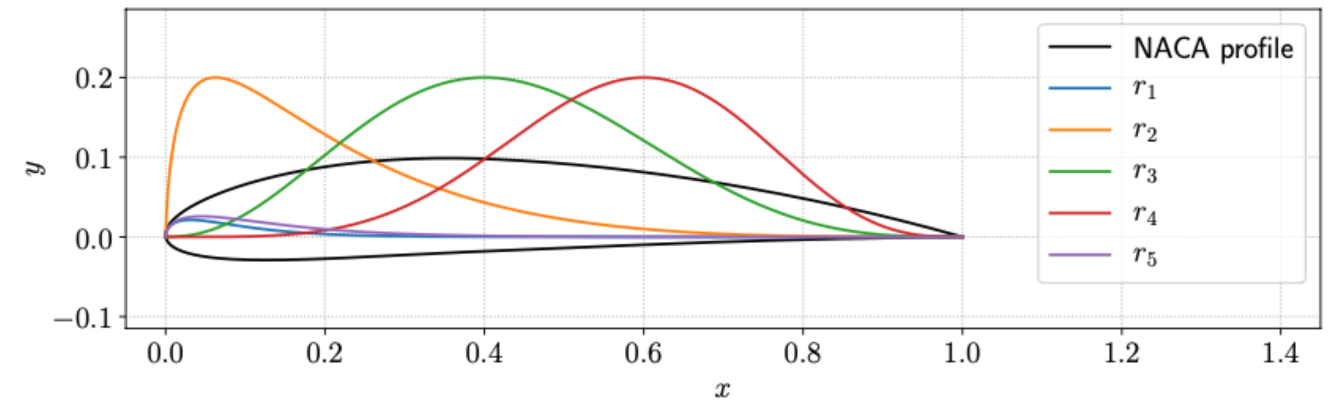


[*] Zhang, Zhang, Hinkle. *Learning nonlinear level sets for dimensionality reduction in function approximation*. Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada.

#Optimization #ASGA

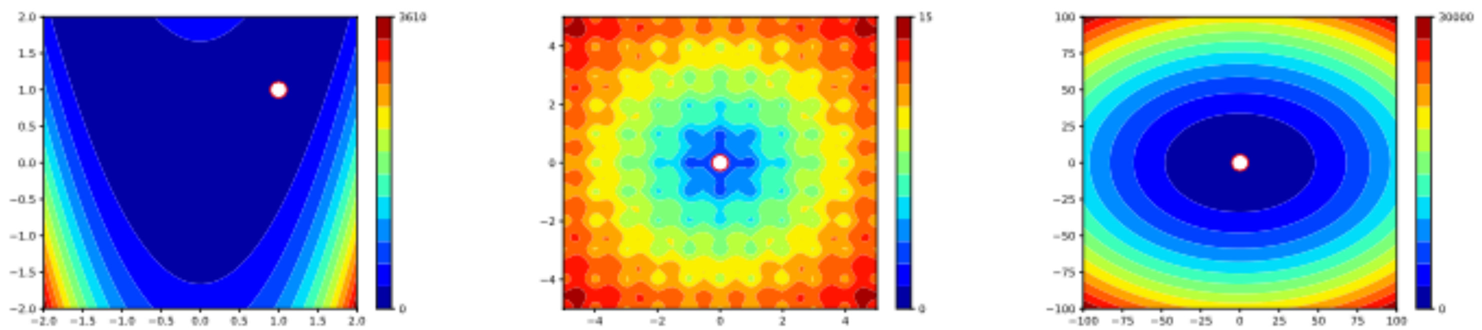
AS meets Genetic Algorithm (ASGA)
New method for high dimensional scalar function optimization



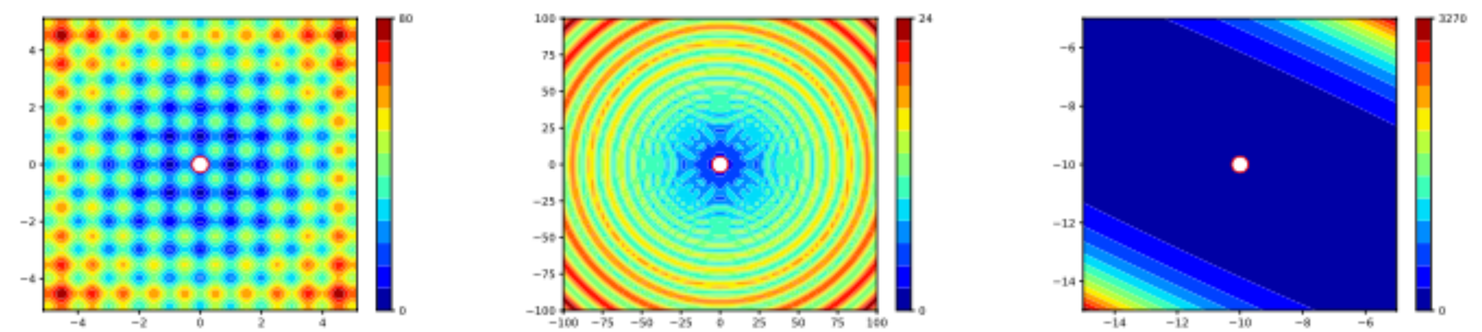


Active Subspaces meets Genetic Algorithm (ASGA) [3]

We developed an extension of the genetic algorithm (GA) which exploits the active subspaces (AS) property to evolve the individuals on a lower dimensional space. In many cases, GA requires in fact more function evaluations than others optimization method to converge to the optimum. Thus, complex and **high-dimensional functions may result intractable** with the standard algorithm. To address this issue, we propose to linearly map the input parameter space of the original function onto its AS before the evolution, performing the *mutation* and *mate* processes in a lower dimensional space.



(a) Rosenbrock function. (b) Ackley function. (c) Bohachevsky function.



(d) Rastrigin function. (e) Schaffer N. 7 function. (f) Zakharov function.

[3] N. Demo, M. Tezzele, and G. Rozza. "*A supervised learning approach involving active subspaces for an efficient genetic algorithm in high-dimensional optimization problems*". Submitted, 2020. arXiv:2006.07282

Recast of the optimization problem

We address the constrained global optimization problem of a real-valued continuous function, in the context of genetic algorithms, defined as

$$\min_{\boldsymbol{\mu} \in \Omega \subset \mathbb{R}^k} f(\boldsymbol{\mu})$$

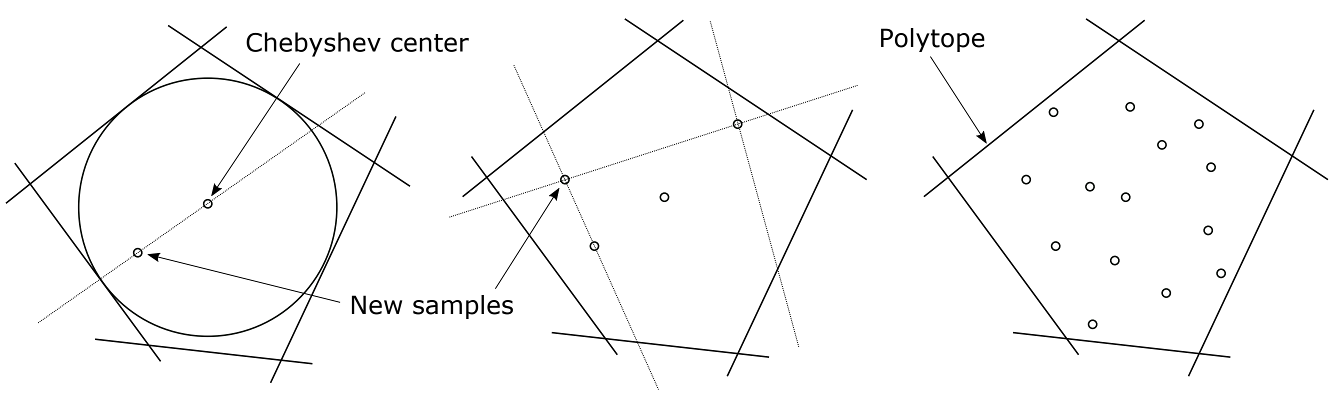
To fight the curse of dimensionality and speed up the convergence we exploit the AS of the target function to **select the best individuals in the reduced parameter space**, mutate and mate them, and successively to map them in the full parameter space. We recast the optimization problem for each generation of individuals over the **polytope** defined below

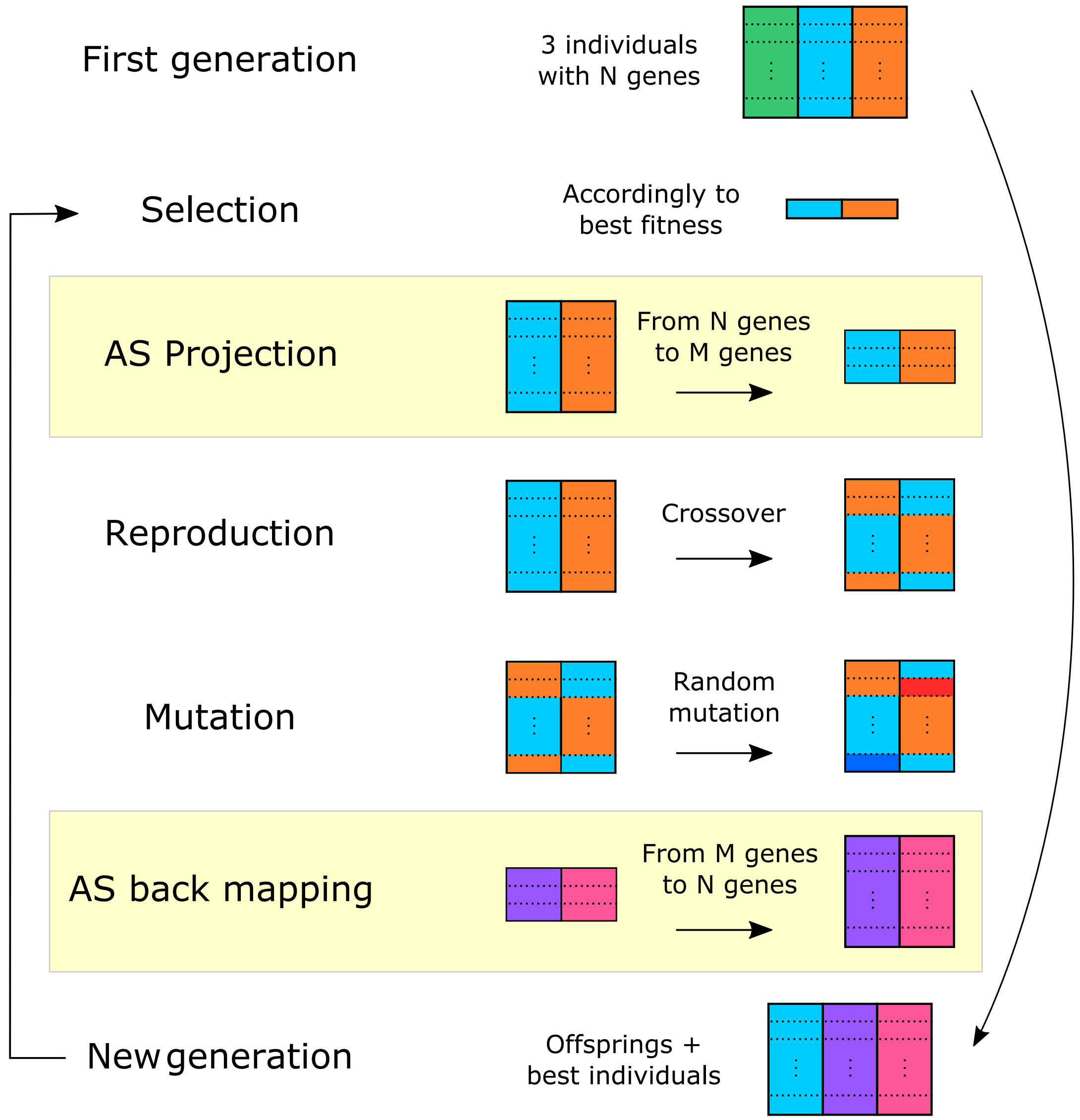
$$\min_{\substack{\boldsymbol{\mu}_M \in \mathcal{P} \subset \mathbb{R}^M \\ \boldsymbol{\mu} \in \Omega}} g(\boldsymbol{\mu}_M = \mathbf{W}_1^T \boldsymbol{\mu})$$

$$\mathcal{P} := \{\boldsymbol{\mu}_M = \mathbf{W}_1^T \boldsymbol{\mu} \mid \boldsymbol{\mu} \in \Omega\}$$

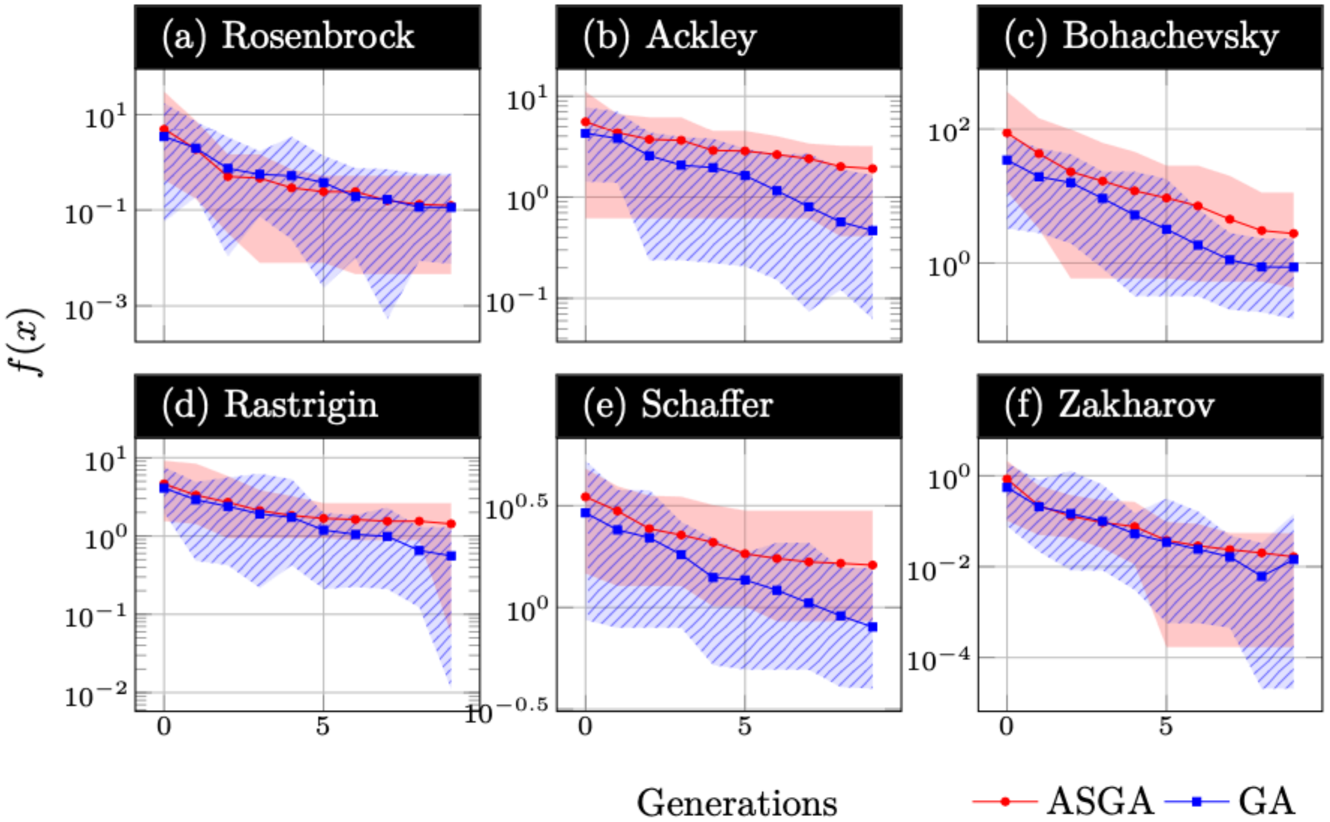
$$g(\mathbf{y}) := f(\mathbf{x}_{\mathbf{y}}) \qquad \forall \mathbf{y} \in \mathcal{Y},$$

$$\mathcal{Y} := \{\mathbf{y} \in \mathcal{P} \mid \exists \mathbf{x}_{\mathbf{y}} \in \Omega \text{ s.t. } \mathbf{y} = \mathbf{W}_1^T \mathbf{x}_{\mathbf{y}}\}.$$





ASGA for 2D parameter spaces: a comparison

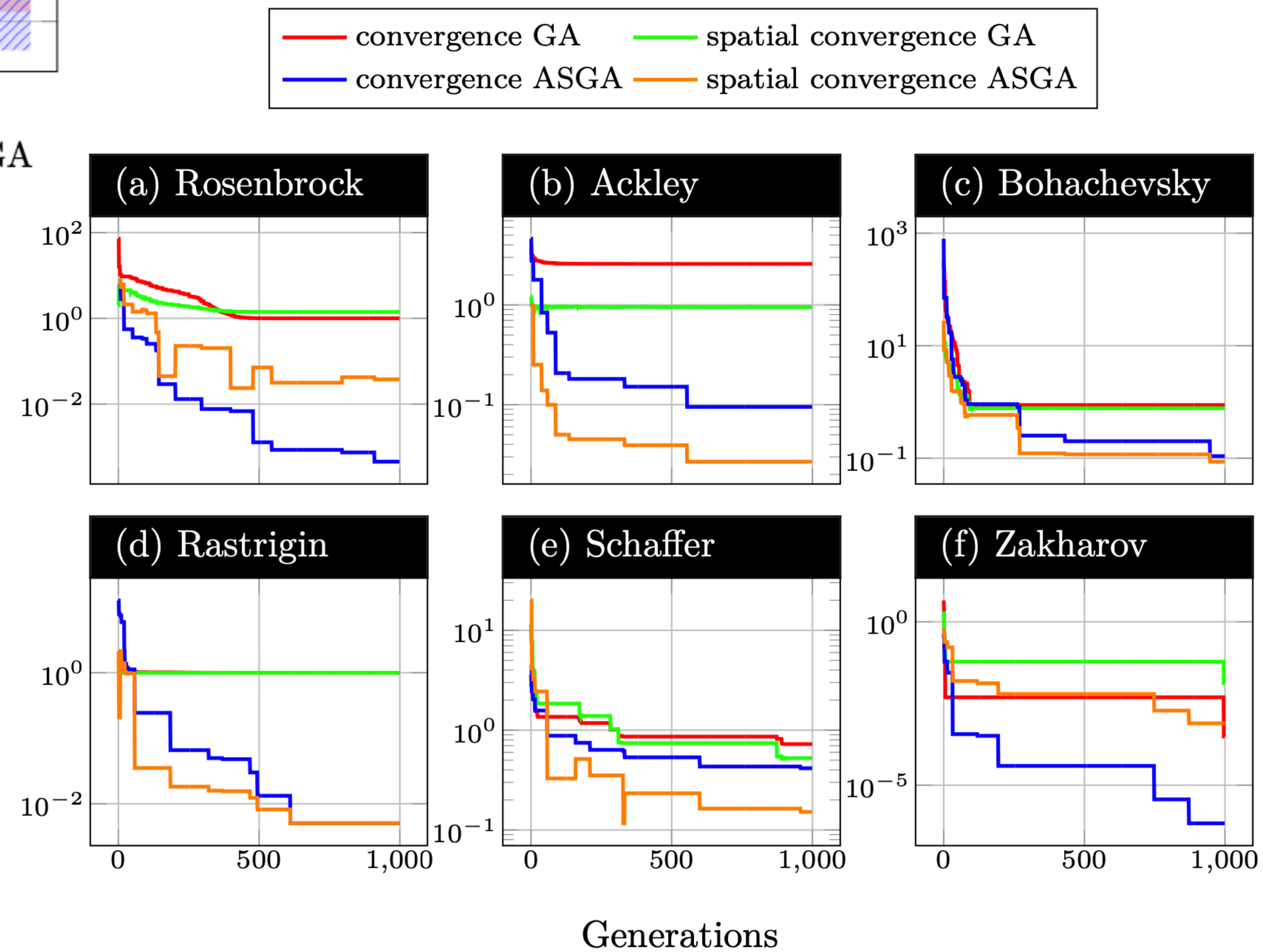


The plots show the results over 15 runs for each benchmark function. The shaded areas show the interval between minimum and maximum.

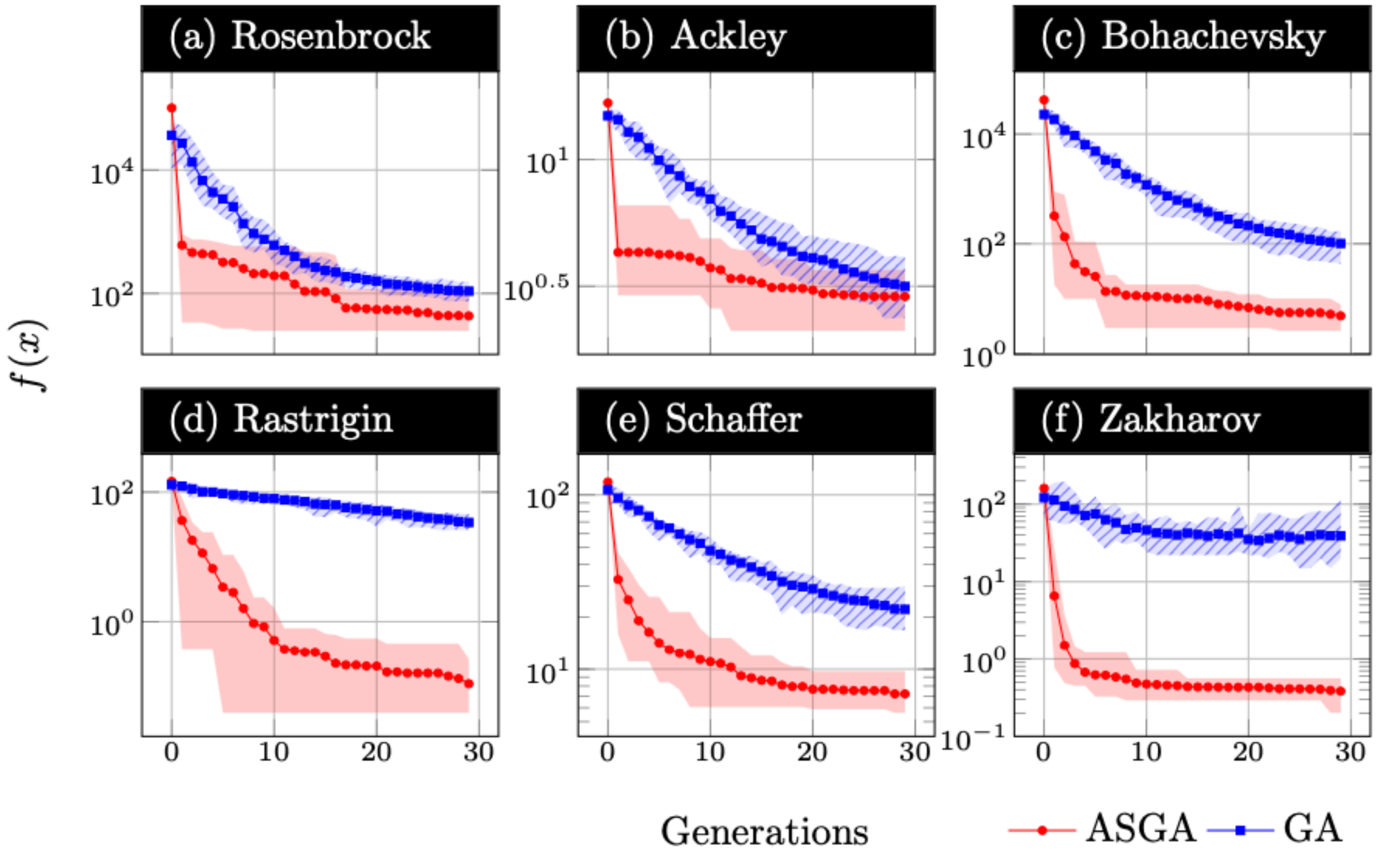
The AS dimension is fixed to 1 for all the tests.

For such low dimensional spaces the gain is minimal with respect to the convergence of the target function, at least for the first generations.

If we have a look at the spatial convergence of x we can see a clear gain.

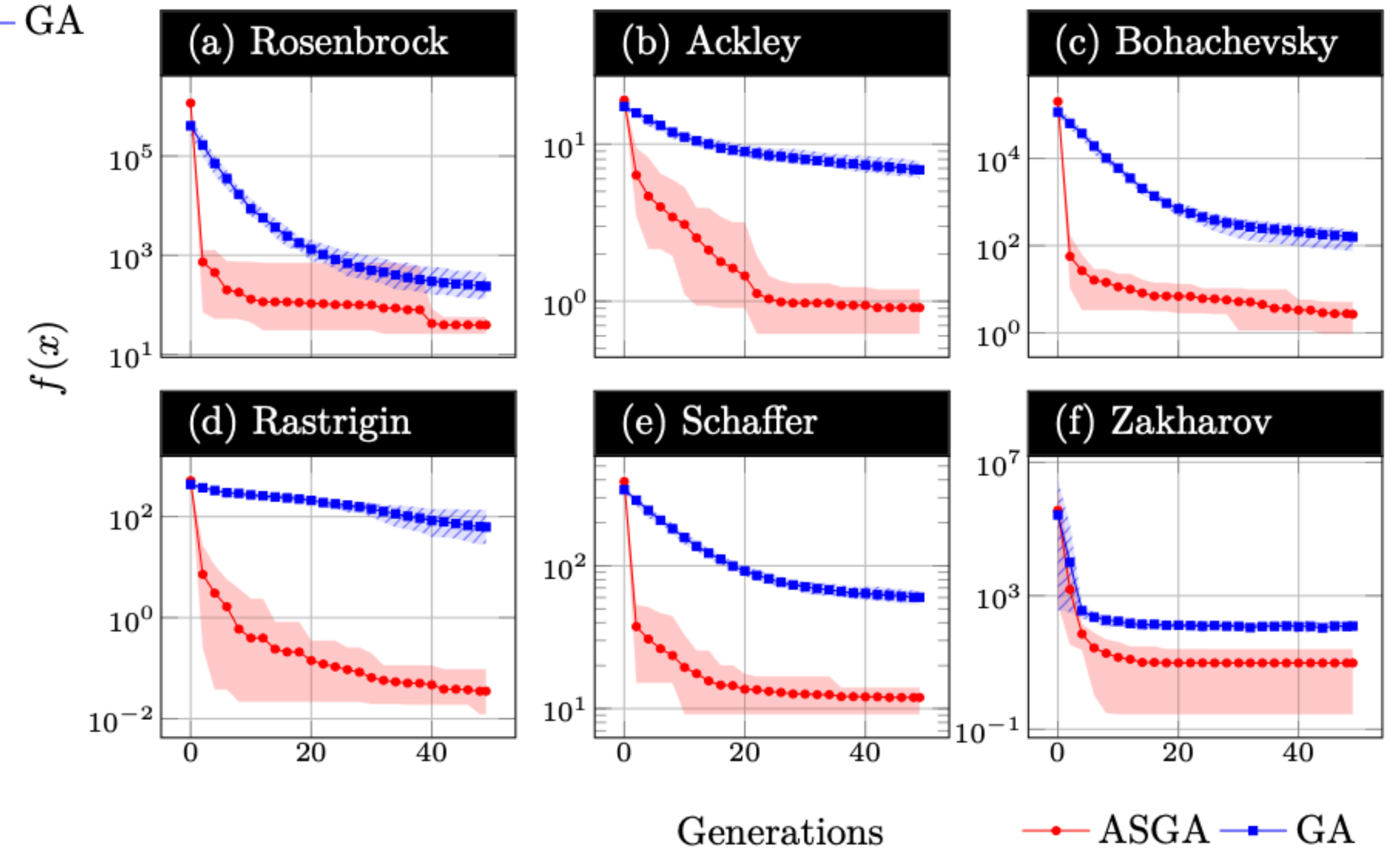


ASGA for 15D and 40D parameter spaces: a comparison



Input dimension = 15
Number of generations = 30
Successive runs = 15

Input dimension = 40
Number of generations = 50
Successive runs = 15

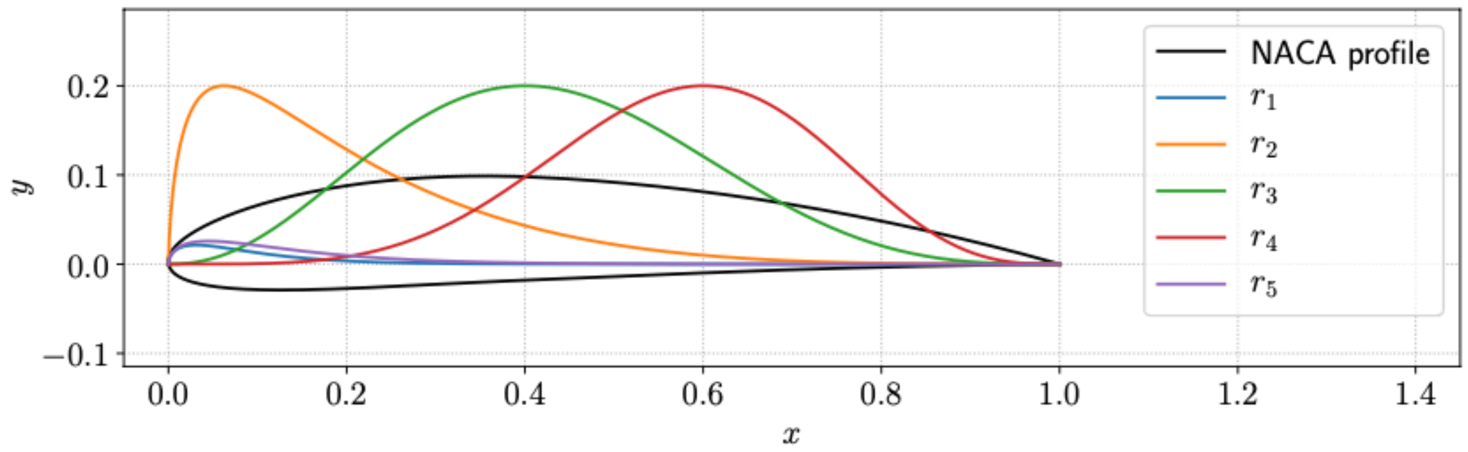


The gain is remarkable and it increases as we increase the dimension of the input space.

N. Demo, M. Tezzele, and G. Rozza. "*A supervised learning approach involving active subspaces for an efficient genetic algorithm in high-dimensional optimization problems*". Submitted, 2020. arXiv:2006.07282

ASGA for an aeronautical shape optimization problem

NACA 4412 morphing



$$y^+ = \overline{y^+} + \sum_{i=1}^{5} a_i r_i,$$

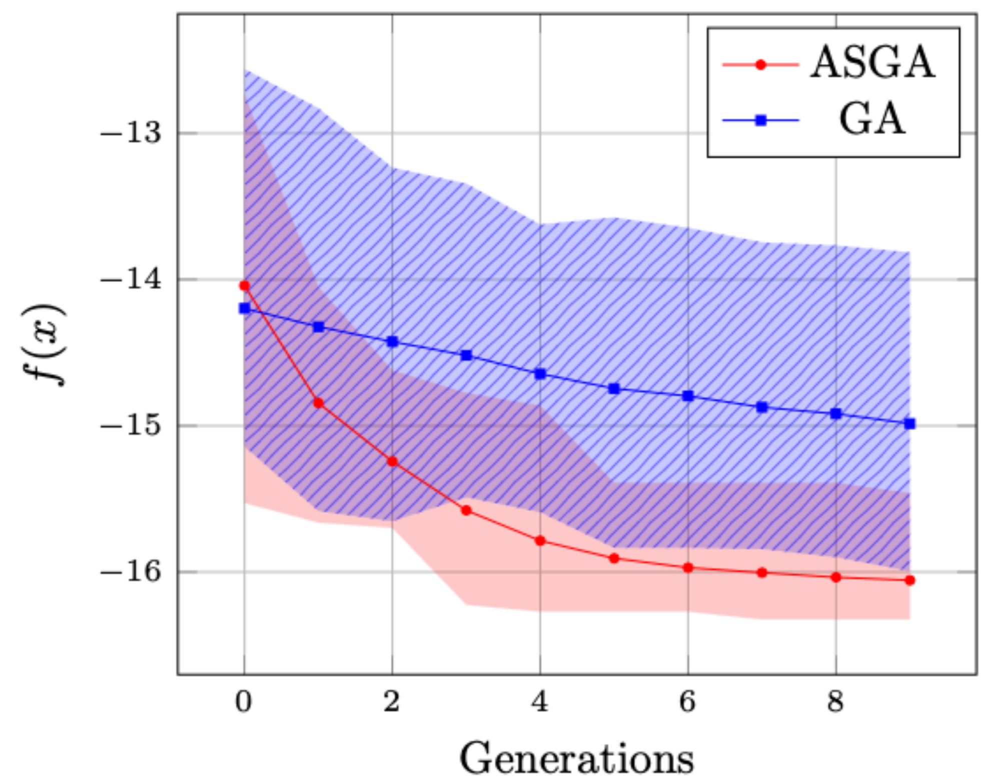
$$y^+ = \overline{y^-} - \sum_{i=1}^{5} b_i r_i,$$

We deform a NACA 4412 airfoil profile with **10 parameters**, which are the coefficients **a**, and **b**, used to morph the upper and lower part of the profile. Further details can be found in [4].

The objective function to minimize is the opposite of the **lift-to-drag coefficient**.

On the right the comparison of the standard GA and the ASGA.

Lift-to-drag coefficient optimization



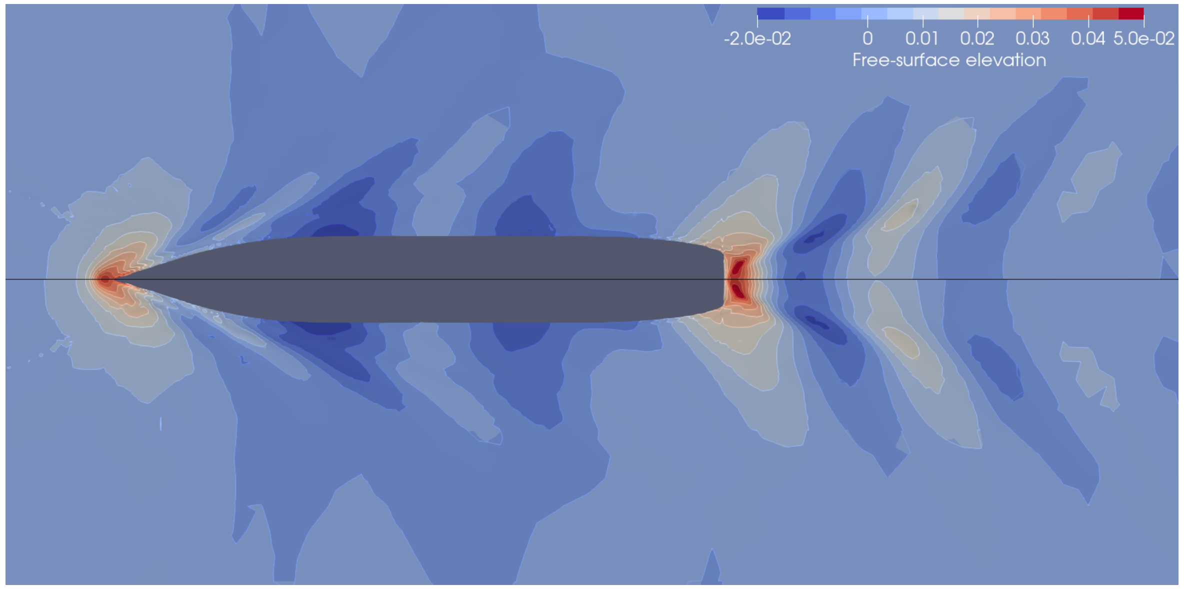
[4] M. Tezzele, N. Demo, G. Stabile, A. Mola, and G. Rozza. "*Enhancing CFD predictions in shape design problems by model and parameter space reduction*". AMSES 7(1) 40, 2020.

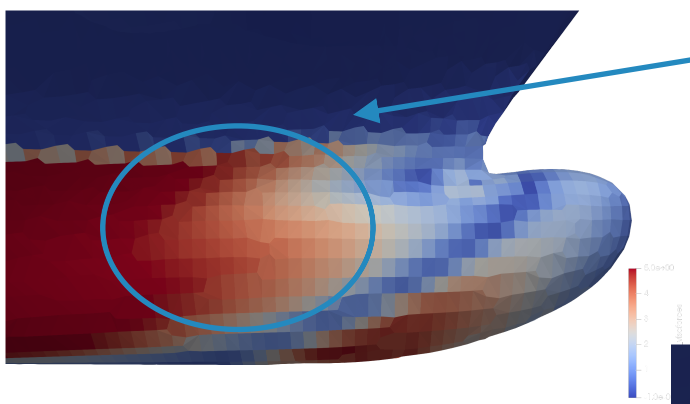
ASGA for a naval engineering optimization problem [5]

We minimize the total resistance coefficient which depends on 10 free form deformation parameters

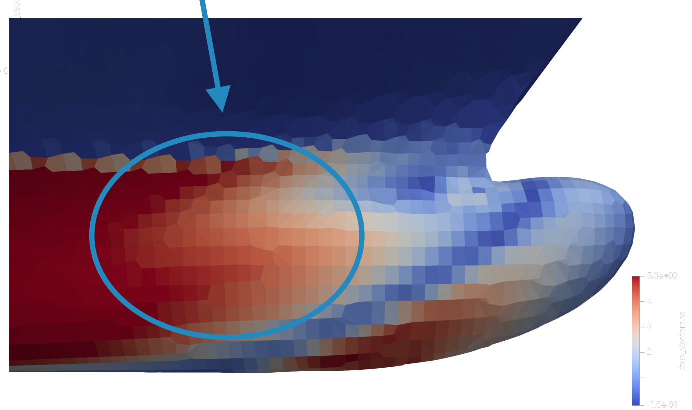
$$\min_{\boldsymbol{\mu}} C_t \equiv \min_{\boldsymbol{\mu}} \int_{\Omega(\boldsymbol{\mu})} \frac{\tau_x \rho - p n_x}{\frac{1}{2}\rho V^2 S}$$

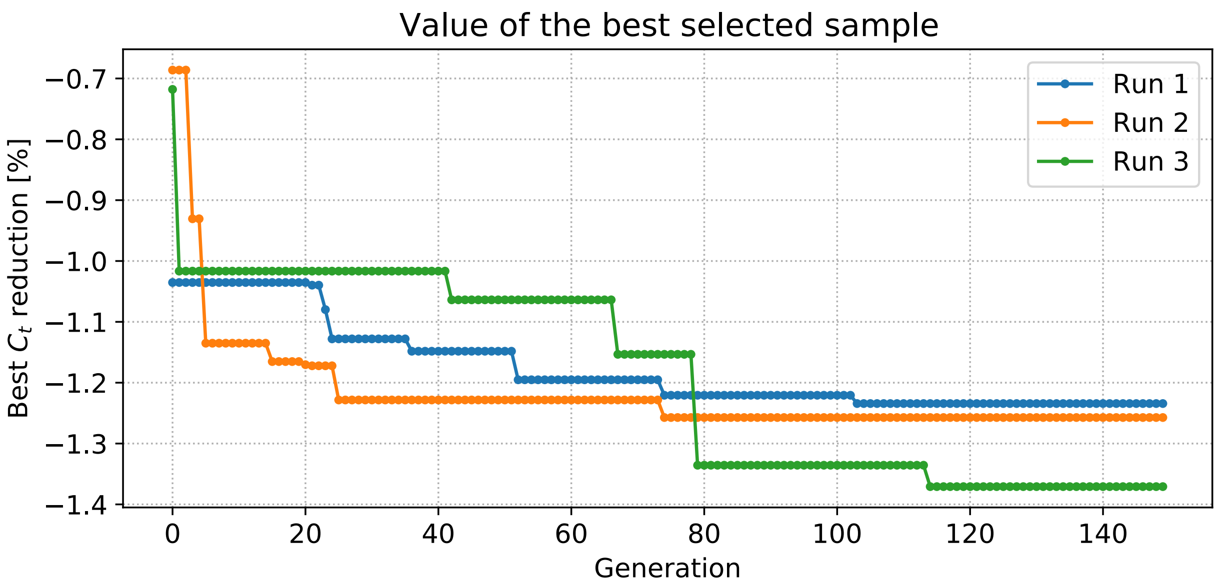
Original free surface elevation (top) and optimized (bottom)





The optimal hull presents lower pressures downstream of the bulbous bow





[5] N. Demo, M. Tezzele, A. Mola, and G. Rozza. "*Hull shape design optimization with parameter space and model reductions, and self-learning mesh morphing*". Journal of Marine Science and Engineering, 9(2):185, 2021

Final considerations

Conclusions:

- We presented a new **kernel-based extension of active subspaces** for both scalar and vectorial output functions. And we tested it on a CFD application.
- We proposed a new modification of the classical genetic algorithm method based on active subspaces, called **ASGA method**. We proved its capabilities on benchmark test functions and on a CFD shape optimization problem.
- Check out the new Python package called **ATHENA** to enhance numerical analysis with parameter space reduction techniques! It is available at github.com/mathLab/ATHENA.

Future works:

- A **multi-fidelity** approach involving AS (preliminary work in [6]) and KAS.
- A **multi-objective** extension of the ASGA method.
- The use of **Reversible Artificial Neural Networks** for parameter space reduction.

[6] F. Romor, M. Tezzele, and G. Rozza. "*Multi-fidelity data fusion for the approximation of scalar functions with low intrinsic dimensionality through active subspaces*". Proceedings in Applied Mathematics & Mechanics, 2020. arXiv:2010.08349