# ACTIVE SUBSPACES for dimension reduction in parameter studies

**PAUL CONSTANTINE**
Ben L. Fryrear Assistant Professor
Applied Mathematics & Statistics
Colorado School of Mines

activesubspaces.org
@DrPaulynomial

In collaboration with:

**RACHEL WARD**
**UT Austin**

**ARMIN EFTEKHARI**
**UT Austin**

This material is based upon work supported by the U.S. Department of Energy Office of Science, Office of Advanced Scientific Computing Research, Applied Mathematics program under Award Number DE-SC-0011077.

**SLIDES:** goo.gl/cK6goL

**DISCLAIMER:** These slides are meant to complement the oral presentation. Use out of context at your own risk.

# TAKE-HOMES

An active subspace is a type of low-dimensional structure in a function of several variables.

# TAKE-HOMES

An active subspace is a type of low-dimensional structure in a function of several variables.

We have tools for identifying and exploiting active subspaces for parameter studies.

# TAKE-HOMES

An active subspace is a type of low-dimensional structure in a function of several variables.

We have tools for identifying and exploiting active subspaces for parameter studies.

Active subspaces appear in a wide range of physical models.

# TAKE-HOMES

An active subspace is a type of low-dimensional structure in a function of several variables.

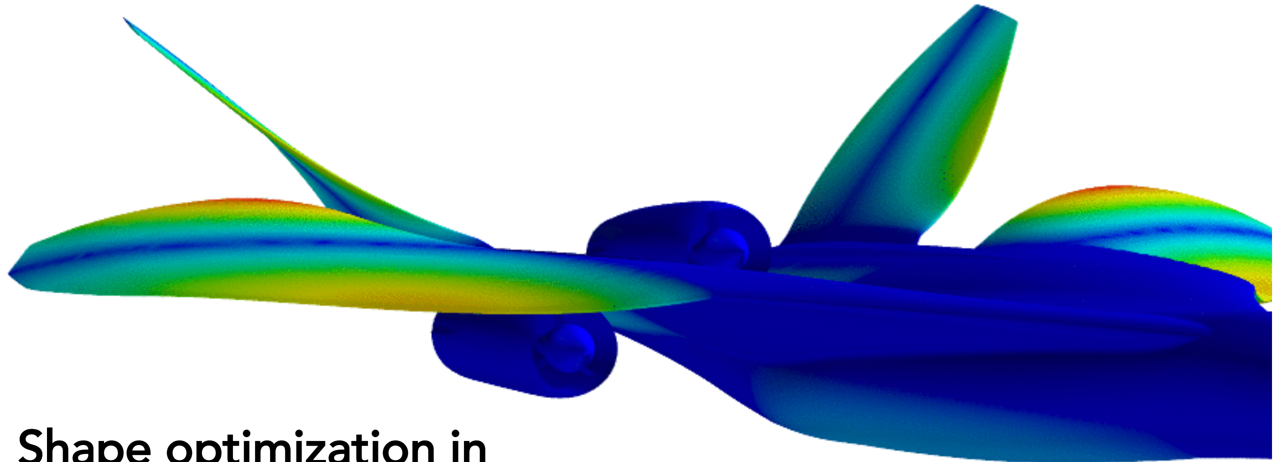
We have tools for identifying and exploiting active subspaces for parameter studies.

Active subspaces appear in a wide range of physical models.

Active subspaces are closely related to ridge approximation.

$$f(\mathbf{x})$$

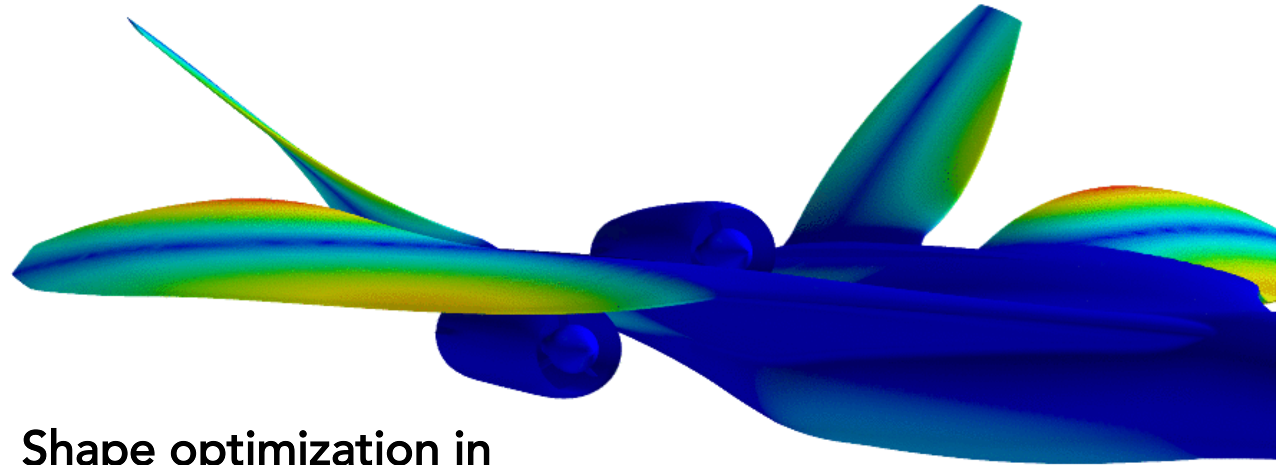
$$f(\mathbf{x})$$



Shape optimization in aerospace vehicles

*(with J. Alonso, T. Lukaczyk)*

$$f(\mathbf{x})$$

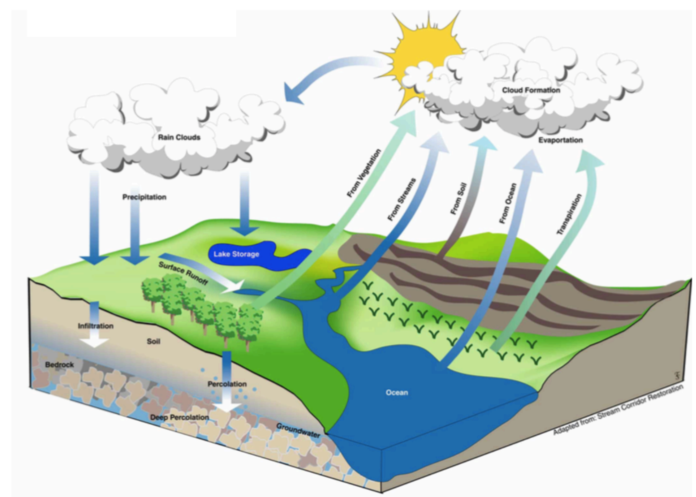


**Shape optimization in aerospace vehicles**

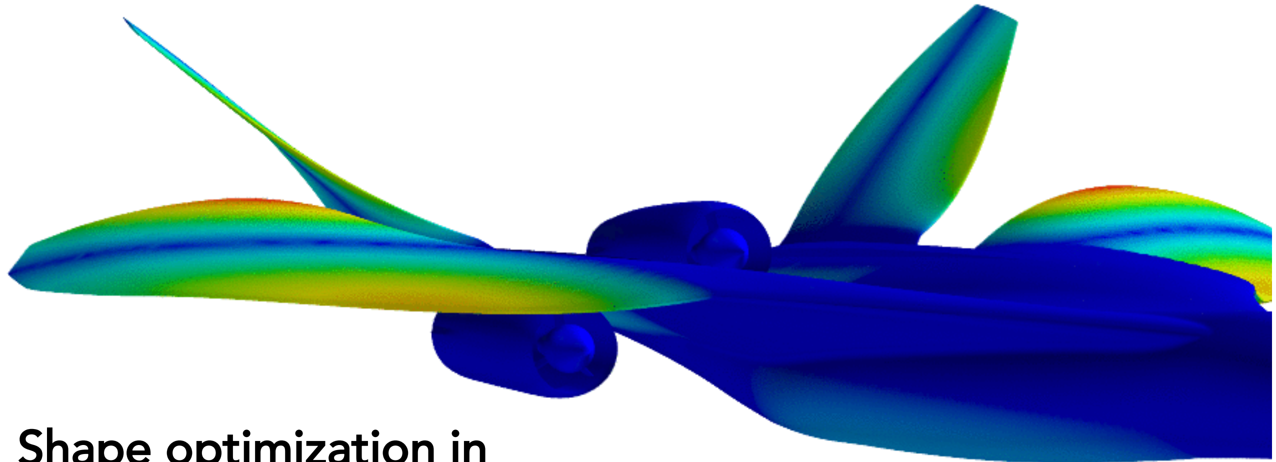
*(with J. Alonso, T. Lukaczyk)*

**Sensitivity analysis in integrated hydrologic models**

*(with R. Maxwell, J. Jefferson, J. Gilbert)*

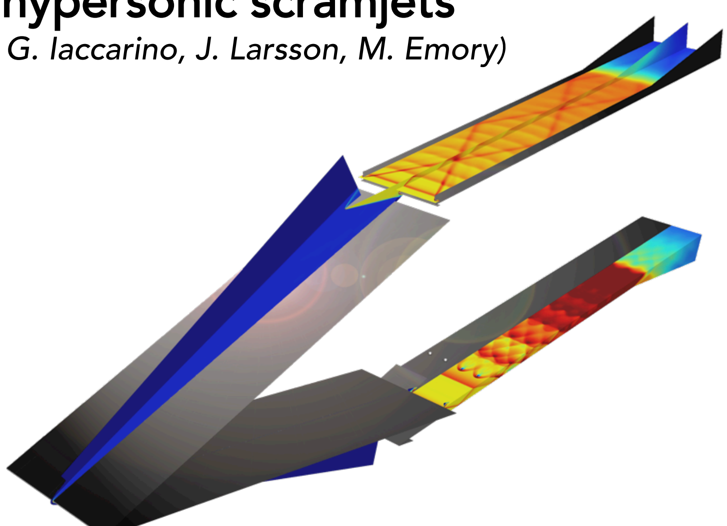


$$f(\mathbf{x})$$



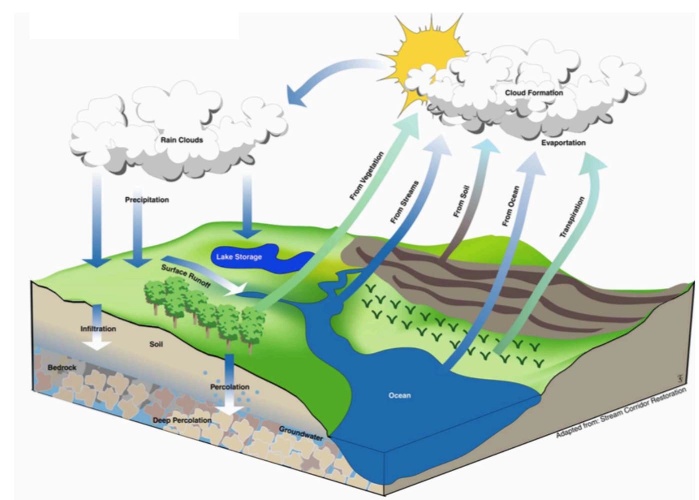
**Shape optimization in aerospace vehicles**
*(with J. Alonso, T. Lukaczyk)*

**Uncertainty quantification for hypersonic scramjets**
*(with G. Iaccarino, J. Larsson, M. Emory)*

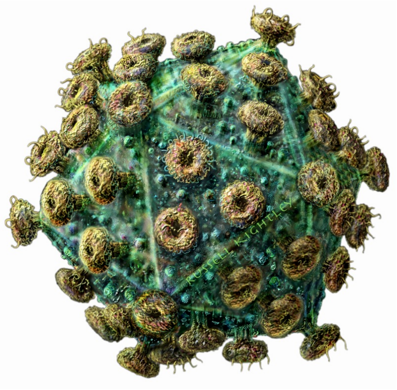


**Sensitivity analysis in integrated hydrologic models**

*(with R. Maxwell, J. Jefferson, J. Gilbert)*



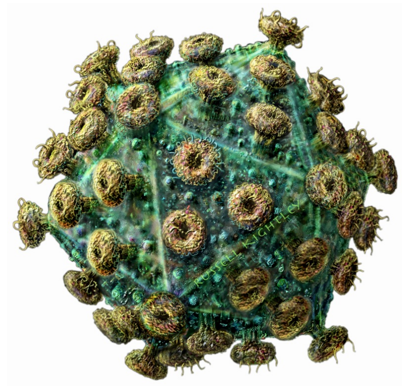
$f(\mathbf{x})$



**Sensitivity analysis in HIV modeling**

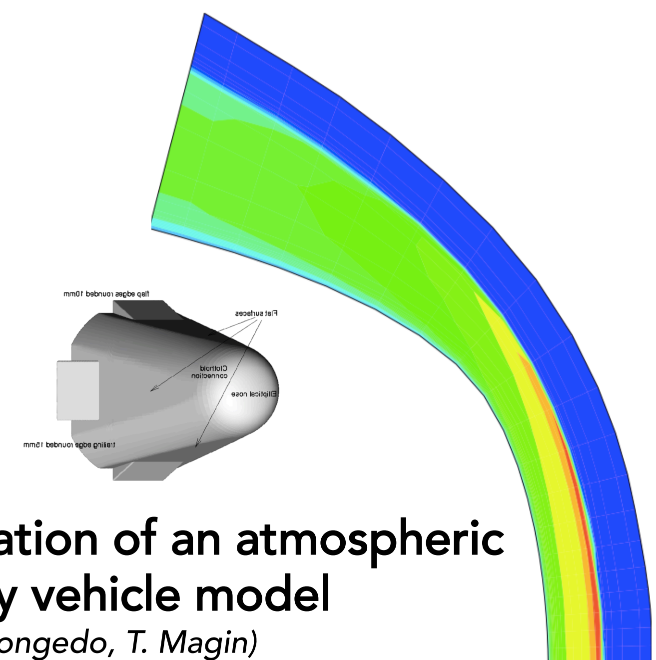
*(with T. Loudon, S. Pankavich)*

$$f(\mathbf{x})$$



**Sensitivity analysis in HIV modeling**

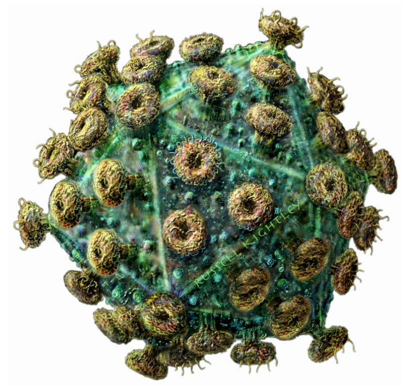
*(with T. Loudon, S. Pankavich)*



**Calibration of an atmospheric reentry vehicle model**

*(with P. Congedo, T. Magin)*

$$f(\mathbf{x})$$

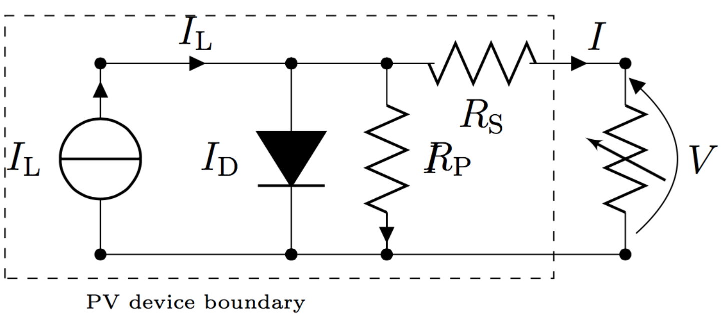


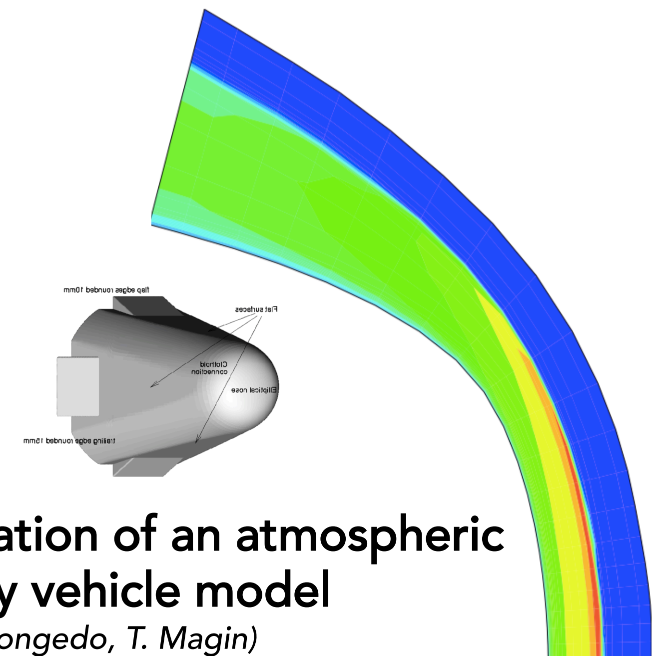
**Sensitivity analysis in HIV modeling**

*(with T. Loudon, S. Pankavich)*

**Sensitivity analysis in solar cell models**

*(with B. Zaharatos, M. Campanelli)*

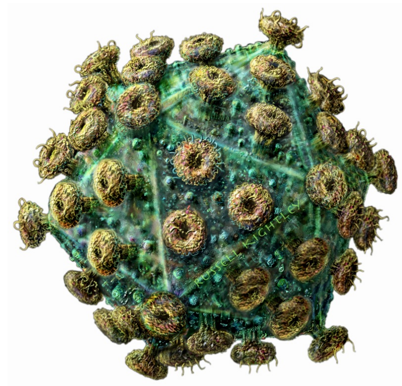


**Calibration of an atmospheric reentry vehicle model**

*(with P. Congedo, T. Magin)*

$$f(\mathbf{x})$$

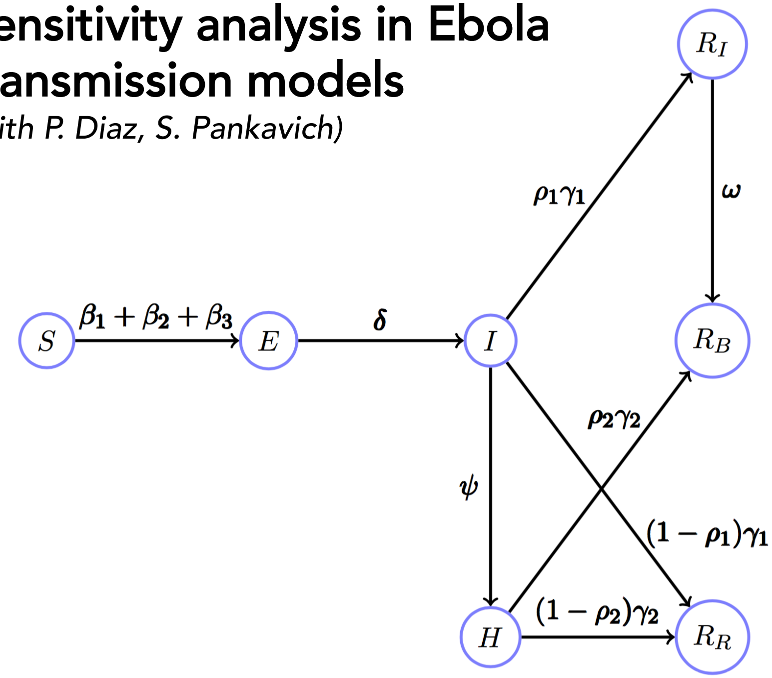


## Sensitivity analysis in HIV modeling

*(with T. Loudon, S. Pankavich)*

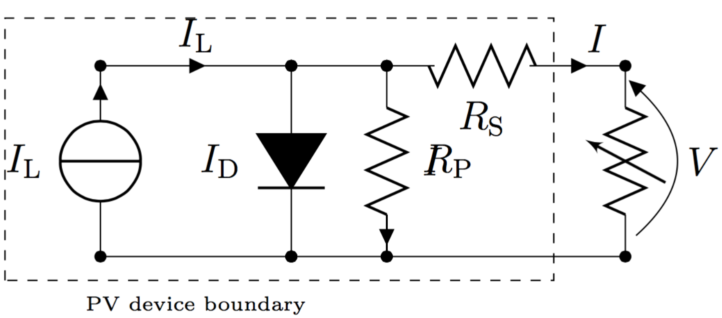
## Sensitivity analysis in Ebola transmission models

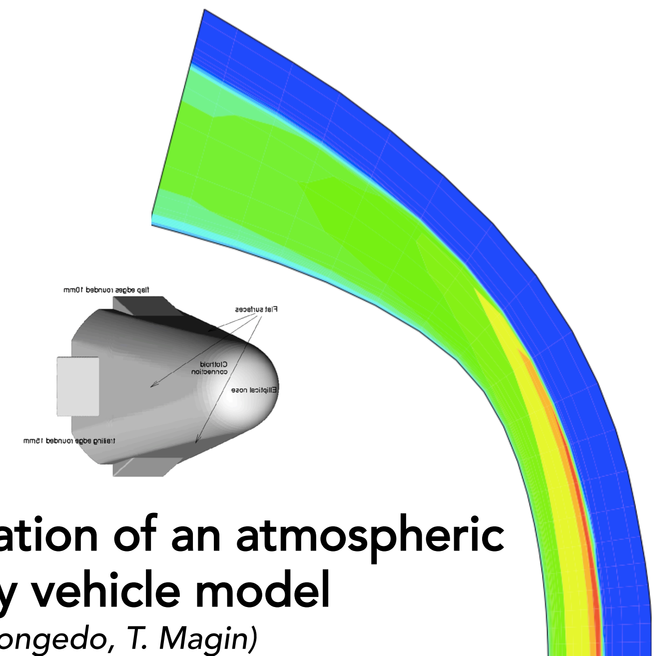
*(with P. Diaz, S. Pankavich)*




## Sensitivity analysis in solar cell models

*(with B. Zaharatos, M. Campanelli)*



## Calibration of an atmospheric reentry vehicle model

*(with P. Congedo, T. Magin)*

$$f(\mathbf{x})$$

## Lithium ion battery model

*(with A. Doostan)*

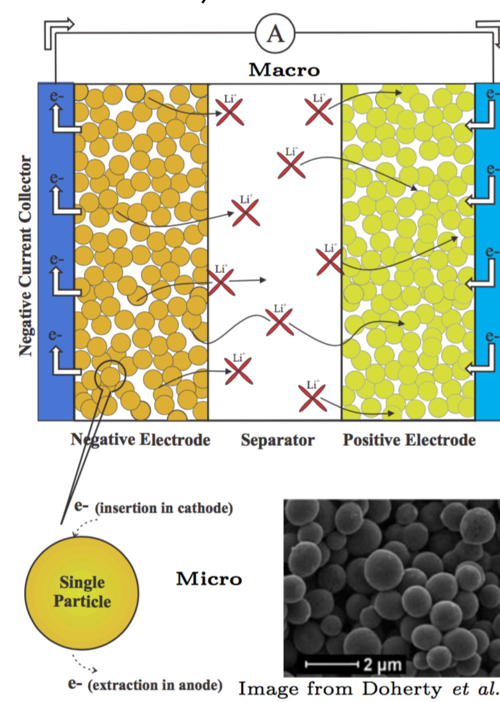
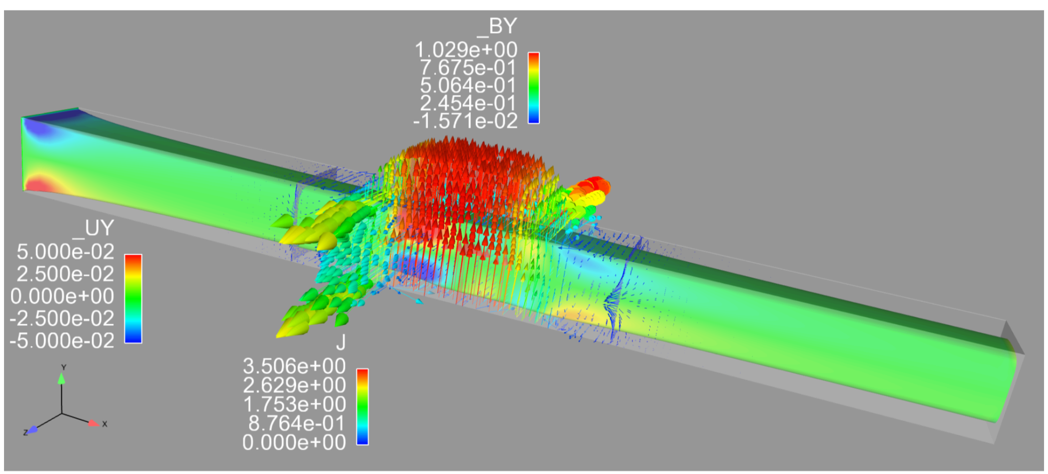



Image from Doherty *et al.* (2010)

$$f(\mathbf{x})$$

## Magnetohydrodynamics generator model

*(with A. Glaws, T. Wildey, J. Shadid)*




## Lithium ion battery model

*(with A. Doostan)*

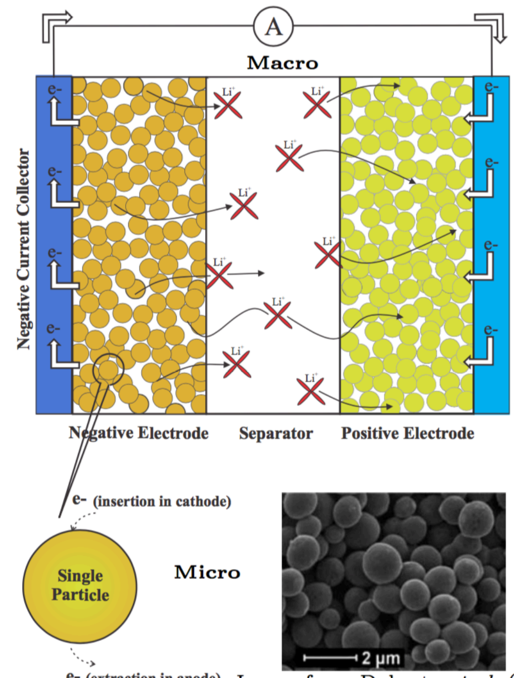
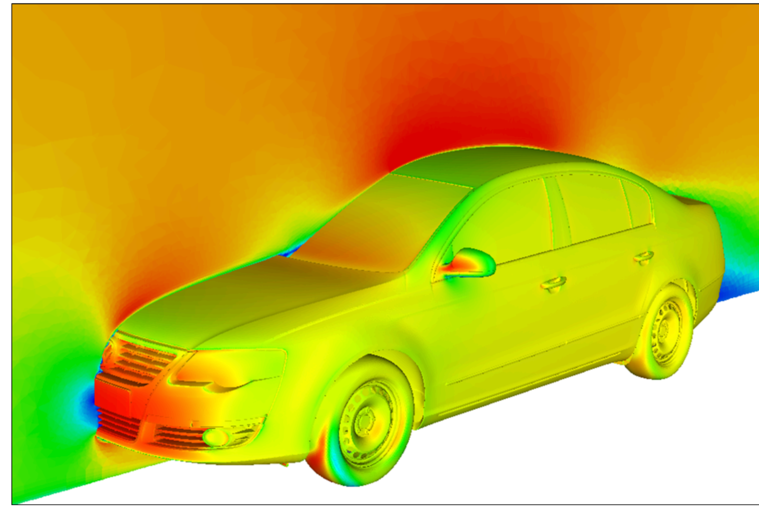



Image from Doherty *et al.* (2010)

$$f(\mathbf{x})$$

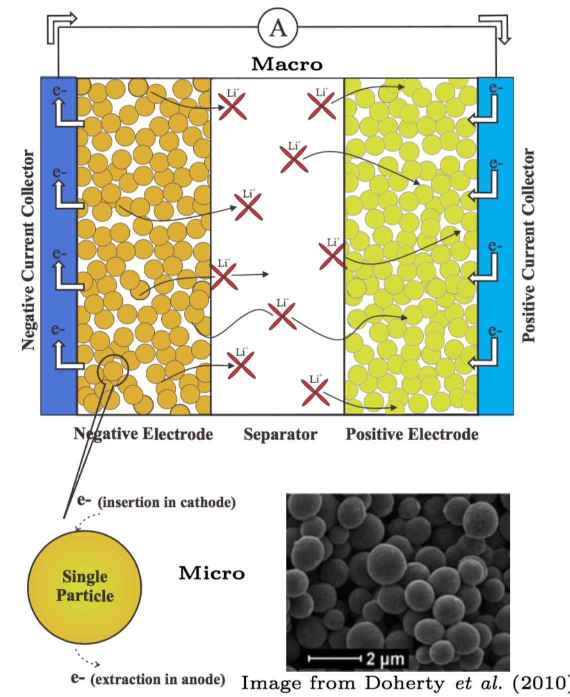
## Vehicle design

*(with C. Othmer, J. Alonso)*



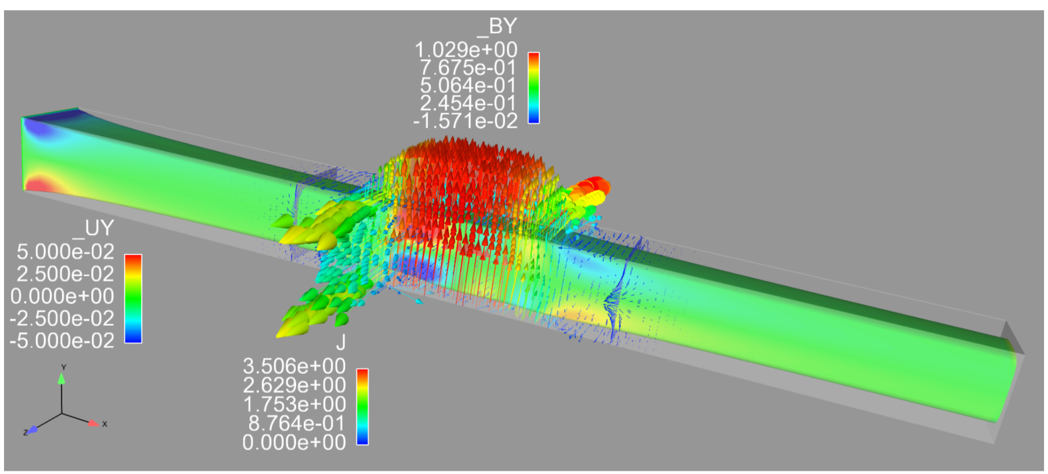
## Lithium ion battery model

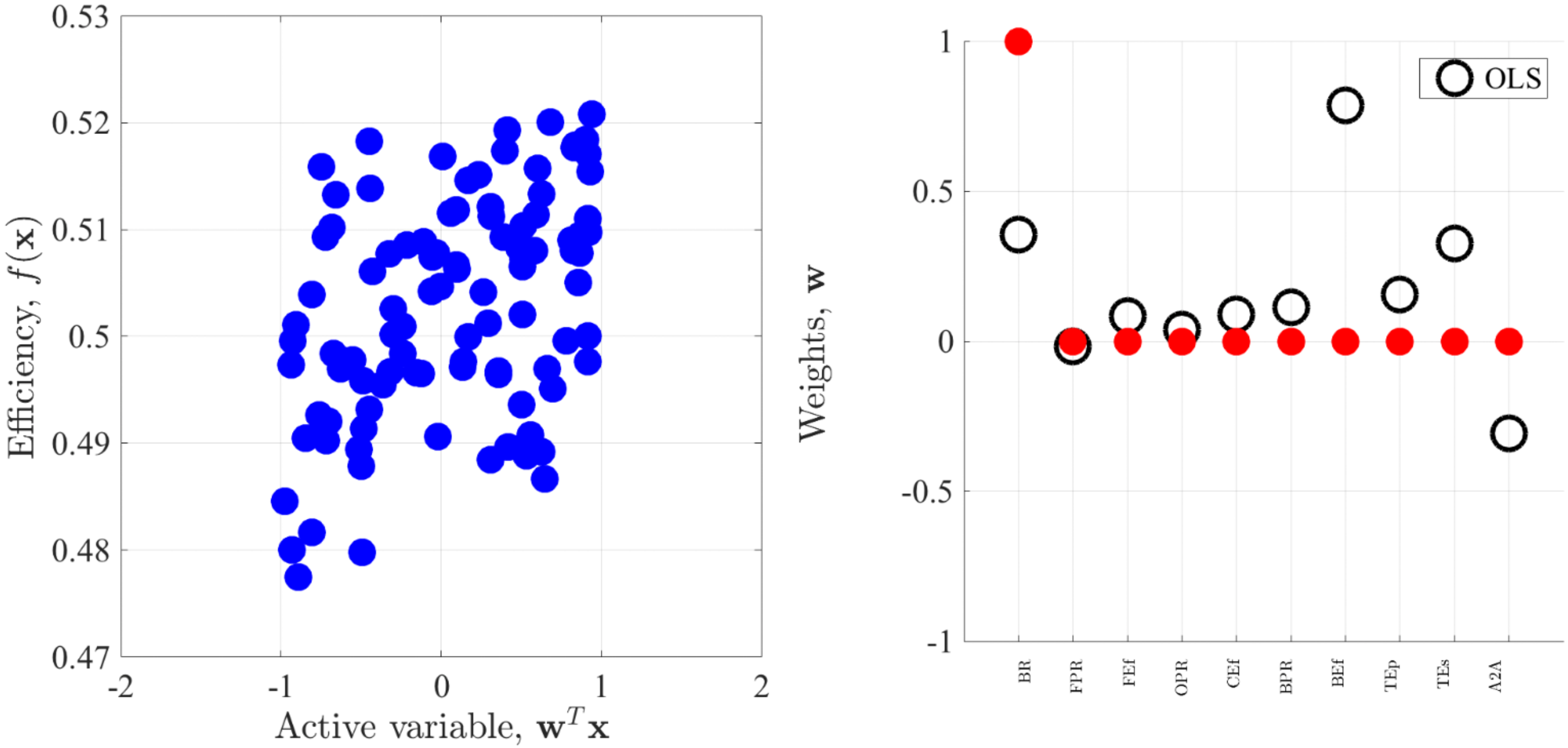
*(with A. Doostan)*




## Magnetohydrodynamics generator model

*(with A. Glaws, T. Wildey, J. Shadid)*





0.53
0.52
0.51
0.5
0.49
0.48
0.47
Efficiency, $f(\mathbf{x})$
-2
-1
0
1
2
Active variable, $\mathbf{w}^T\mathbf{x}$
1
0.5
0
-0.5
-1
Weights, $\mathbf{w}$
OLS
BR
FPR
FEf
OPR
CEf
BPR
BEf
TEp
TEs
A2A

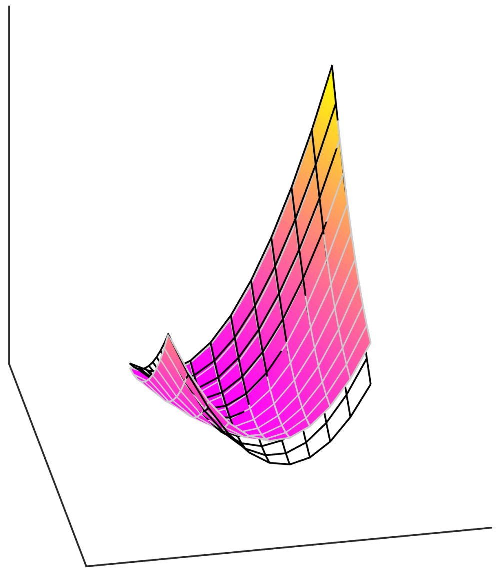
# Ridge approximation

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T \mathbf{x})$$

where

$$\boldsymbol{U}^T : \mathbb{R}^m \to \mathbb{R}^n$$

$$g : \mathbb{R}^n \to \mathbb{R}$$



# Ridge approximation

What is U?

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T\mathbf{x})$$

# Ridge approximation

What is U?

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T\mathbf{x})$$

What is g?

# Ridge approximation

What is the
approximation error?

What is U?

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T\mathbf{x})$$

What is g?

# Ridge approximation

What is the
approximation error?

$$f(\mathbf{x}) \boxed{\approx} g(\boldsymbol{U}^T\mathbf{x})$$

# Ridge approximation

What is the
approximation error?

$$f(\mathbf{x}) \boxed{\approx} g(\boldsymbol{U}^T\mathbf{x})$$

Use the weighted root-mean-squared error:

$$\left\| f(\mathbf{x}) - g(\boldsymbol{U}^T\mathbf{x}) \right\|_{L^2(\rho)} = \left( \int \left(f(\mathbf{x}) - g(\boldsymbol{U}^T\mathbf{x})\right)^2 \rho(\mathbf{x})\, d\mathbf{x} \right)^{\frac{1}{2}}$$

# Ridge approximation

What is the approximation error?

$$f(\mathbf{x}) \boxed{\approx} g(\boldsymbol{U}^T\mathbf{x})$$

Use the weighted root-mean-squared error:

$$\left\| f(\mathbf{x}) - g(\boldsymbol{U}^T\mathbf{x}) \right\|_{L^2(\rho)} = \left( \int (f(\mathbf{x}) - g(\boldsymbol{U}^T\mathbf{x}))^2 \boxed{\rho(\mathbf{x})} d\mathbf{x} \right)^{\frac{1}{2}}$$

Given weight function

# Ridge approximation

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T \mathbf{x})$$

What is g?

# Ridge approximation

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T\mathbf{x})$$

What is g?

Use the conditional average:

$$\mu(\mathbf{y}) = \int f(\boldsymbol{U}\mathbf{y} + \boldsymbol{V}\mathbf{z})\,\pi(\mathbf{z}|\mathbf{y})\,d\mathbf{z}$$

# Ridge approximation

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T\mathbf{x})$$

What is g?

Use the conditional average:

$$\mu(\mathbf{y}) = \int f(\boldsymbol{U}\mathbf{y} + \boldsymbol{V}\mathbf{z})\, \pi(\mathbf{z}|\mathbf{y})\, d\mathbf{z}$$

Subspace coordinates

# Ridge approximation

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T\mathbf{x})$$

What is g?

Use the conditional average:

$$\mu(\mathbf{y}) = \int f(\boldsymbol{U}\mathbf{y} + \boldsymbol{V}\mathbf{z})\, \pi(\mathbf{z}|\mathbf{y})\, d\mathbf{z}$$

Subspace coordinates

Complement subspace and coordinates

# Ridge approximation

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T\mathbf{x})$$

What is g?

Use the conditional average:

Conditional density

$$\mu(\mathbf{y}) = \int f(\boldsymbol{U}\mathbf{y} + \boldsymbol{V}\mathbf{z})\,\pi(\mathbf{z}|\mathbf{y})\,d\mathbf{z}$$

Subspace coordinates

Complement subspace and coordinates

# Ridge approximation

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T\mathbf{x})$$

What is g?

Use the conditional average:

Conditional density

$$\mu(\mathbf{y}) = \int f(\boldsymbol{U}\mathbf{y} + \boldsymbol{V}\mathbf{z})\,\pi(\mathbf{z}|\mathbf{y})\,d\mathbf{z}$$

Subspace coordinates

Complement subspace and coordinates

$\mu(\boldsymbol{U}^T\mathbf{x})$ is the **best approximation** (Pinkus, 2015).

# Ridge approximation

What is U?

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T \mathbf{x})$$

# Ridge approximation

What is U?

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T \mathbf{x})$$

Define the error function:

$$R(\boldsymbol{U}) = \frac{1}{2} \int (f(\mathbf{x}) - \mu(\boldsymbol{U}^T \mathbf{x}))^2 \, \rho(\mathbf{x}) \, d\mathbf{x}$$

# Ridge approximation

What is U?

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T \mathbf{x})$$

Define the error function:

$$R(\boldsymbol{U}) = \frac{1}{2} \int (f(\mathbf{x}) - \mu(\boldsymbol{U}^T \mathbf{x}))^2 \, \rho(\mathbf{x}) \, d\mathbf{x}$$

Minimize the error:

$$\underset{\boldsymbol{U}}{\text{minimize}} \;\; R(\boldsymbol{U}) \quad \text{subject to} \;\; \boldsymbol{U} \in \mathbb{G}(n, m)$$

# Ridge approximation

What is U?

$$f(\mathbf{x}) \approx g(\boldsymbol{U}^T\mathbf{x})$$

Define the error function:

$$R(\boldsymbol{U}) = \frac{1}{2}\int (f(\mathbf{x}) - \mu(\boldsymbol{U}^T\mathbf{x}))^2\,\rho(\mathbf{x})\,d\mathbf{x}$$

Minimize the error:

$$\underset{\boldsymbol{U}}{\text{minimize}}\ \ R(\boldsymbol{U}) \quad \text{subject to}\ \ \boldsymbol{U} \in \mathbb{G}(n,m)$$

Grassmann manifold of
n-dimensional subspaces

# Define the active subspace

Consider a function and its gradient vector,

$$f = f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m, \quad \nabla f(\mathbf{x}) \in \mathbb{R}^m, \quad \rho : \mathbb{R}^m \rightarrow \mathbb{R}_+$$

# Define the active subspace

Consider a function and its gradient vector,

$$f = f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m, \quad \nabla f(\mathbf{x}) \in \mathbb{R}^m, \quad \rho : \mathbb{R}^m \to \mathbb{R}_+$$

The average outer product of the gradient and its eigendecomposition,

$$\boldsymbol{C} \;=\; \int \nabla f(\mathbf{x})\, \nabla f(\mathbf{x})^T \,\rho(\mathbf{x})\, d\mathbf{x} \;=\; \boldsymbol{W} \Lambda \boldsymbol{W}^T$$

# Define the active subspace

Consider a function and its gradient vector,

$$f = f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m, \quad \nabla f(\mathbf{x}) \in \mathbb{R}^m, \quad \rho : \mathbb{R}^m \to \mathbb{R}_+$$

The average outer product of the gradient and its eigendecomposition,

$$\boldsymbol{C} \;=\; \int \nabla f(\mathbf{x})\, \nabla f(\mathbf{x})^T \rho(\mathbf{x})\, d\mathbf{x} \;=\; \boldsymbol{W} \Lambda \boldsymbol{W}^T$$

Partition the eigendecomposition,

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix}, \qquad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{W}_2 \end{bmatrix}, \qquad \boldsymbol{W}_1 \in \mathbb{R}^{m \times n}$$

# Define the active subspace

Consider a function and its gradient vector,

$$f = f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m, \quad \nabla f(\mathbf{x}) \in \mathbb{R}^m, \quad \rho : \mathbb{R}^m \to \mathbb{R}_+$$

The average outer product of the gradient and its eigendecomposition,

$$\boldsymbol{C} \;=\; \int \nabla f(\mathbf{x})\, \nabla f(\mathbf{x})^T \rho(\mathbf{x})\, d\mathbf{x} \;=\; \boldsymbol{W} \Lambda \boldsymbol{W}^T$$

Partition the eigendecomposition,

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix}, \qquad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{W}_2 \end{bmatrix}, \qquad \boldsymbol{W}_1 \in \mathbb{R}^{m \times n}$$

Rotate and separate the coordinates,

$$\mathbf{x} \;=\; \boldsymbol{W}\boldsymbol{W}^T\mathbf{x} \;=\; \boldsymbol{W}_1\boldsymbol{W}_1^T\mathbf{x} + \boldsymbol{W}_2\boldsymbol{W}_2^T\mathbf{x} \;=\; \boldsymbol{W}_1\mathbf{y} + \boldsymbol{W}_2\mathbf{z}$$

# Define the active subspace

Consider a function and its gradient vector,

$$f = f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m, \quad \nabla f(\mathbf{x}) \in \mathbb{R}^m, \quad \rho : \mathbb{R}^m \rightarrow \mathbb{R}_+$$

The average outer product of the gradient and its eigendecomposition,

$$\boldsymbol{C} \;=\; \int \nabla f(\mathbf{x}) \, \nabla f(\mathbf{x})^T \, \rho(\mathbf{x}) \, d\mathbf{x} \;=\; \boldsymbol{W} \Lambda \boldsymbol{W}^T$$

Partition the eigendecomposition,

$$\Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix}, \qquad \boldsymbol{W} = \begin{bmatrix} \boldsymbol{W}_1 & \boldsymbol{W}_2 \end{bmatrix}, \qquad \boldsymbol{W}_1 \in \mathbb{R}^{m \times n}$$

Rotate and separate the coordinates,

active variables

inactive variables

$$\mathbf{x} \;=\; \boldsymbol{W}\boldsymbol{W}^T\mathbf{x} \;=\; \boldsymbol{W}_1\boldsymbol{W}_1^T\mathbf{x} + \boldsymbol{W}_2\boldsymbol{W}_2^T\mathbf{x} \;=\; \boldsymbol{W}_1\mathbf{y} + \boldsymbol{W}_2\mathbf{z}$$

# The eigenpairs identify perturbations that change the function more, on average.

**LEMMA**

$$\lambda_i = \int \left(\mathbf{w}_i^T \nabla f(\mathbf{x})\right)^2 \rho(\mathbf{x})\, d\mathbf{x}, \qquad i = 1, \ldots, m$$

# The eigenpairs identify perturbations that change the function more, on average.

**LEMMA**

$$\lambda_i \;=\; \int \left(\mathbf{w}_i^T \, \nabla f(\mathbf{x})\right)^2 \rho(\mathbf{x}) \, d\mathbf{x}, \qquad i = 1, \ldots, m$$

**LEMMA**

$$\int \|\nabla_{\mathbf{y}} f(\mathbf{x})\|_2^2 \, \rho(\mathbf{x}) \, d\mathbf{x} = \lambda_1 + \cdots + \lambda_n$$

$$\int \|\nabla_{\mathbf{z}} f(\mathbf{x})\|_2^2 \, \rho(\mathbf{x}) \, d\mathbf{x} = \lambda_{n+1} + \cdots + \lambda_m$$

# An approximation result

$$\left\| f(\mathbf{x}) - \mu(\boldsymbol{W}_1^T \mathbf{x}) \right\|_{L^2(\rho)} \;\leq\; C\left(\lambda_{n+1} + \cdots + \lambda_m\right)^{\frac{1}{2}}$$

# An approximation result

Conditional average

$$\left\| f(\mathbf{x}) - \mu(\boldsymbol{W}_1^T \mathbf{x}) \right\|_{L^2(\rho)} \leq C \left(\lambda_{n+1} + \cdots + \lambda_m\right)^{\frac{1}{2}}$$

# An approximation result

Conditional average

$$\left\| f(\mathbf{x}) - \mu(\boldsymbol{W}_1^T \mathbf{x}) \right\|_{L^2(\rho)} \;\leq\; C \left( \lambda_{n+1} + \cdots + \lambda_m \right)^{\frac{1}{2}}$$

Active subspace

# An approximation result

Conditional average

Poincaré constant

$$\left\| f(\mathbf{x}) - \mu(\boldsymbol{W}_1^T \mathbf{x}) \right\|_{L^2(\rho)} \leq C \left( \lambda_{n+1} + \cdots + \lambda_m \right)^{\frac{1}{2}}$$

Active subspace

# An approximation result

Conditional average

Poincaré constant

$$\left\| f(\mathbf{x}) - \mu(\boldsymbol{W}_1^T \mathbf{x}) \right\|_{L^2(\rho)} \leq C \left( \lambda_{n+1} + \cdots + \lambda_m \right)^{\frac{1}{2}}$$

Active subspace

Eigenvalues associated with inactive subspace

# The active subspace is nearly stationary.

Recall:

$$R(\boldsymbol{U}) \;=\; \frac{1}{2}\int \left(f(\mathbf{x}) - \mu(\boldsymbol{U}^T\mathbf{x})\right)^2 \rho(\mathbf{x})\, d\mathbf{x}$$

Assume (1) Lipschitz continuous function
(2) Gaussian density function

# The active subspace is nearly stationary.

Recall:

$$R(\boldsymbol{U}) \;=\; \frac{1}{2}\int (f(\mathbf{x}) - \mu(\boldsymbol{U}^T\mathbf{x}))^2\,\rho(\mathbf{x})\,d\mathbf{x}$$

Assume (1) Lipschitz continuous function
(2) Gaussian density function

$$\|\bar{\nabla} R(\boldsymbol{W}_1)\|_F \;\leq\; L\left(2m^{\frac{1}{2}} + (m-n)^{\frac{1}{2}}\right)(\lambda_{n+1} + \cdots + \lambda_m)^{\frac{1}{2}}$$

# The active subspace is nearly stationary.

Recall:

$$R(\boldsymbol{U}) \;=\; \frac{1}{2}\int (f(\mathbf{x}) - \mu(\boldsymbol{U}^T\mathbf{x}))^2\,\rho(\mathbf{x})\,d\mathbf{x}$$

Assume (1) Lipschitz continuous function
(2) Gaussian density function

Gradient on the
Grassmann manifold

$$\|\bar{\nabla} R(\boldsymbol{W}_1)\|_F \;\leq\; L\left(2m^{\frac{1}{2}} + (m-n)^{\frac{1}{2}}\right)(\lambda_{n+1} + \cdots + \lambda_m)^{\frac{1}{2}}$$

# The active subspace is nearly stationary.

Recall:

$$R(\boldsymbol{U}) \;=\; \frac{1}{2}\int (f(\mathbf{x}) - \mu(\boldsymbol{U}^T\mathbf{x}))^2\,\rho(\mathbf{x})\,d\mathbf{x}$$

Assume (1) Lipschitz continuous function
(2) Gaussian density function

Gradient on the Grassmann manifold

$$\|\bar{\nabla} R(\boldsymbol{W}_1)\|_F \;\leq\; L\left(2m^{\frac{1}{2}} + (m-n)^{\frac{1}{2}}\right)(\lambda_{n+1} + \cdots + \lambda_m)^{\frac{1}{2}}$$

Active subspace

# The active subspace is nearly stationary.

Recall:

$$R(\boldsymbol{U}) \;=\; \frac{1}{2}\int (f(\mathbf{x}) - \mu(\boldsymbol{U}^T\mathbf{x}))^2\,\rho(\mathbf{x})\,d\mathbf{x}$$

Assume (1) Lipschitz continuous function
(2) Gaussian density function

Gradient on the Grassmann manifold

$$\|\bar{\nabla} R(\boldsymbol{W}_1)\|_F \;\leq\; L\left(2m^{\frac{1}{2}} + (m-n)^{\frac{1}{2}}\right)(\lambda_{n+1} + \cdots + \lambda_m)^{\frac{1}{2}}$$

Active subspace

Frobenius norm

# The active subspace is nearly stationary.

Recall:

$$R(\boldsymbol{U}) \;=\; \frac{1}{2}\int \left(f(\mathbf{x}) - \mu(\boldsymbol{U}^T\mathbf{x})\right)^2 \rho(\mathbf{x})\, d\mathbf{x}$$

Assume (1) Lipschitz continuous function
(2) Gaussian density function

Gradient on the Grassmann manifold

Lipschitz constant

Dimensions

$$\|\bar{\nabla} R(\boldsymbol{W}_1)\|_F \;\leq\; L\left(2m^{\frac{1}{2}} + (m-n)^{\frac{1}{2}}\right)\left(\lambda_{n+1} + \cdots + \lambda_m\right)^{\frac{1}{2}}$$

Active subspace

Frobenius norm

# The active subspace is nearly stationary.

Recall:

$$R(\boldsymbol{U}) \;=\; \frac{1}{2}\int \left(f(\mathbf{x}) - \mu(\boldsymbol{U}^T\mathbf{x})\right)^2 \rho(\mathbf{x})\, d\mathbf{x}$$

Assume (1) Lipschitz continuous function
(2) Gaussian density function

Gradient on the Grassmann manifold

Lipschitz constant

Dimensions

$$\|\bar{\nabla} R(\boldsymbol{W}_1)\|_F \;\le\; L\left(2m^{\frac{1}{2}} + (m-n)^{\frac{1}{2}}\right)\left(\lambda_{n+1} + \cdots + \lambda_m\right)^{\frac{1}{2}}$$

Active subspace

Frobenius norm

Eigenvalues associated with inactive subspace

**IDEA Use active subspace as the starting point for numerical ridge approximation.**

Given an initial subspace $U_0$ and samples $\{\mathbf{x}_i, f(\mathbf{x}_i)\}$

# IDEA Use active subspace as the starting point for numerical ridge approximation.

Given an initial subspace $\boldsymbol{U}_0$ and samples $\{\mathbf{x}_i, f(\mathbf{x}_i)\}$

(1) Compute $\mathbf{y}_i = \boldsymbol{U}_0^T \mathbf{x}_i$

## IDEA Use active subspace as the starting point for numerical ridge approximation.

Given an initial subspace $\boldsymbol{U}_0$ and samples $\{\mathbf{x}_i, f(\mathbf{x}_i)\}$

(1) Compute $\mathbf{y}_i = \boldsymbol{U}_0^T \mathbf{x}_i$

(2) Fit a polynomial $p_N(\mathbf{y}, \theta)$ with the pairs $\{\mathbf{y}_i, f(\mathbf{x}_i)\}$

$$\theta_* \; = \; \underset{\theta}{\operatorname{argmin}} \sum_i \big(f(\mathbf{x}_i) - p(\mathbf{y}_i, \theta)\big)^2$$

# IDEA Use active subspace as the starting point for numerical ridge approximation.

Given an initial subspace $\boldsymbol{U}_0$ and samples $\{\mathbf{x}_i, f(\mathbf{x}_i)\}$

(1) Compute $\mathbf{y}_i = \boldsymbol{U}_0^T \mathbf{x}_i$

(2) Fit a polynomial $p_N(\mathbf{y}, \theta)$ with the pairs $\{\mathbf{y}_i, f(\mathbf{x}_i)\}$

$$\theta_* = \underset{\theta}{\operatorname{argmin}} \sum_i \big(f(\mathbf{x}_i) - p(\mathbf{y}_i, \theta)\big)^2$$

(3) Minimize residual over subpsaces

$$\boldsymbol{U}_* = \underset{\boldsymbol{U} \in \mathbb{G}(n,m)}{\operatorname{argmin}} \sum_i \Big(f(\mathbf{x}_i) - p(\boldsymbol{U}^T \mathbf{x}_i, \theta_*)\Big)^2$$

## IDEA Use active subspace as the starting point for numerical ridge approximation.

Given an initial subspace $\boldsymbol{U}_0$ and samples $\{\mathbf{x}_i, f(\mathbf{x}_i)\}$

(1) Compute $\mathbf{y}_i = \boldsymbol{U}_0^T \mathbf{x}_i$

(2) Fit a polynomial $p_N(\mathbf{y}, \theta)$ with the pairs $\{\mathbf{y}_i, f(\mathbf{x}_i)\}$

$$\theta_* = \underset{\theta}{\operatorname{argmin}} \sum_i \big( f(\mathbf{x}_i) - p(\mathbf{y}_i, \theta) \big)^2$$

(3) Minimize residual over subpsaces

$$\boldsymbol{U}_* = \underset{\boldsymbol{U} \in \mathbb{G}(n,m)}{\operatorname{argmin}} \sum_i \Big( f(\mathbf{x}_i) - p(\boldsymbol{U}^T \mathbf{x}_i, \theta_*) \Big)^2$$

(4) Set $\boldsymbol{U}_0 = \boldsymbol{U}_*$ and repeat

# An example where it doesn't work

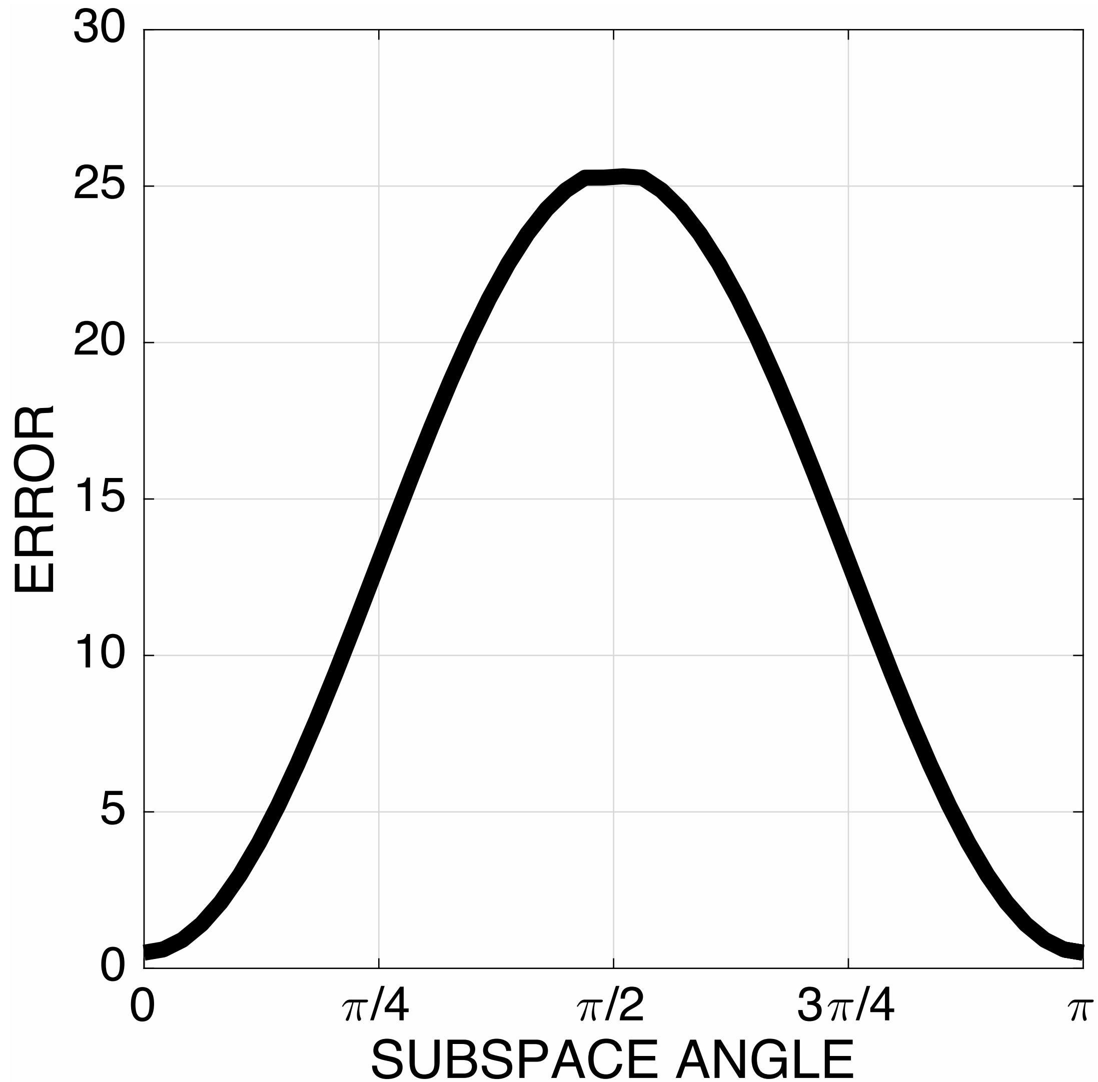
$$f(x_1, x_2) = 5x_1 + \sin(10\pi x_2)$$

$$\boldsymbol{C} = \begin{bmatrix} 25 & 0 \\ 0 & 526 \end{bmatrix}$$

# An example where it doesn't work

$$f(x_1, x_2) = 5x_1 + \sin(10\pi x_2)$$

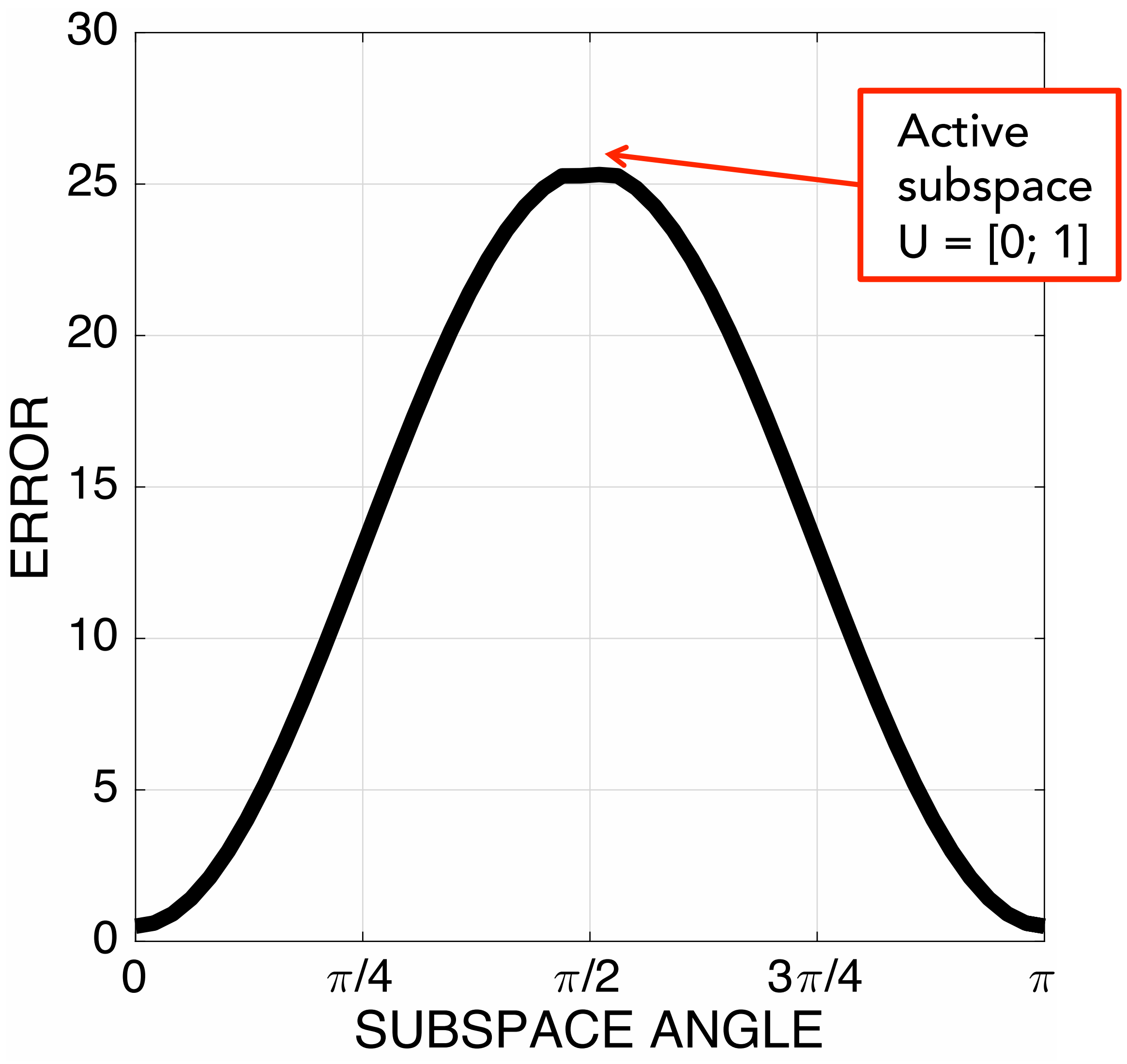
$$\boldsymbol{C} = \begin{bmatrix} 25 & 0 \\ 0 & 526 \end{bmatrix}$$



# An example where it doesn't work

$$f(x_1, x_2) = 5x_1 + \sin(10\pi x_2)$$

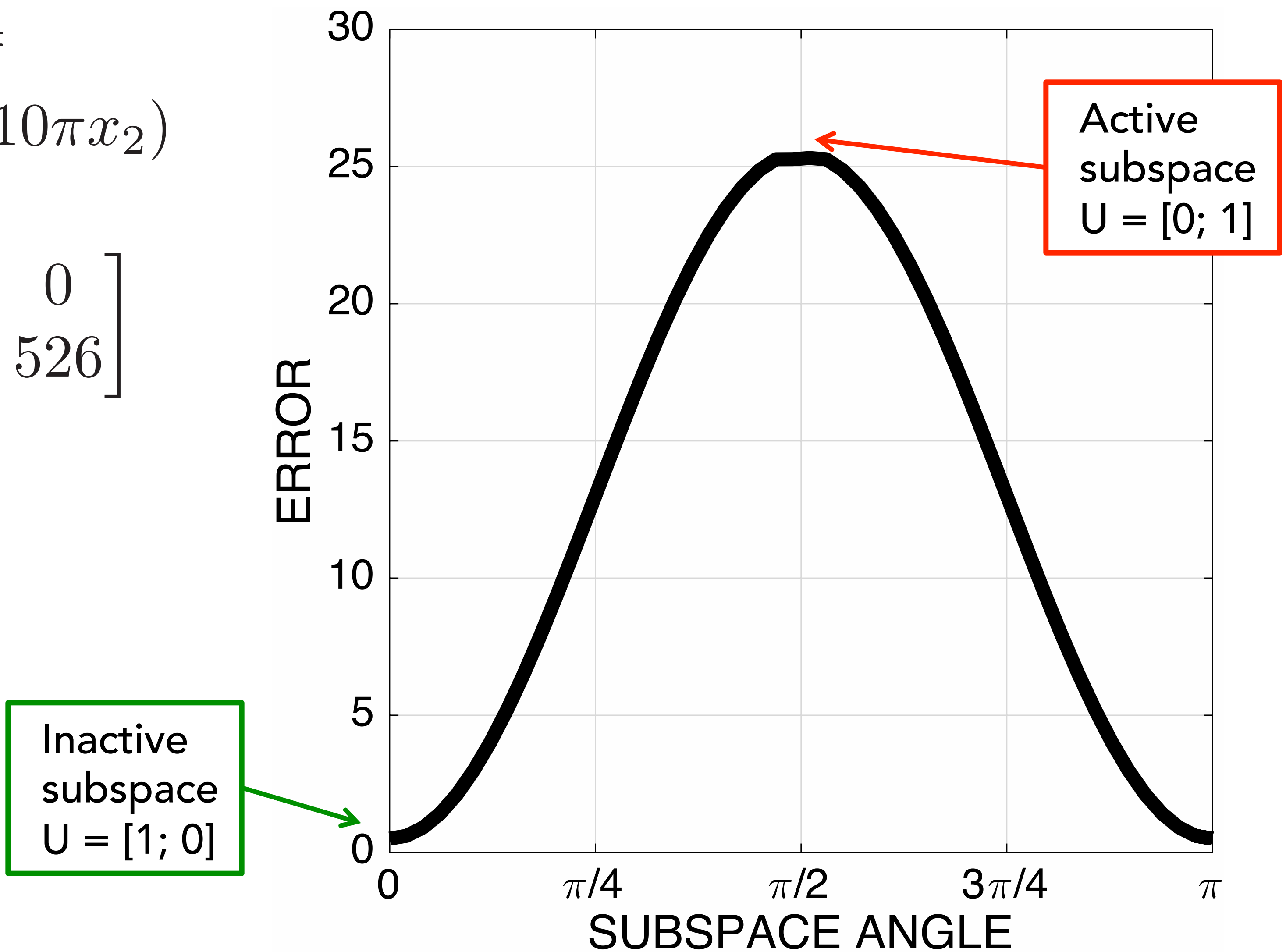
$$\boldsymbol{C} = \begin{bmatrix} 25 & 0 \\ 0 & 526 \end{bmatrix}$$



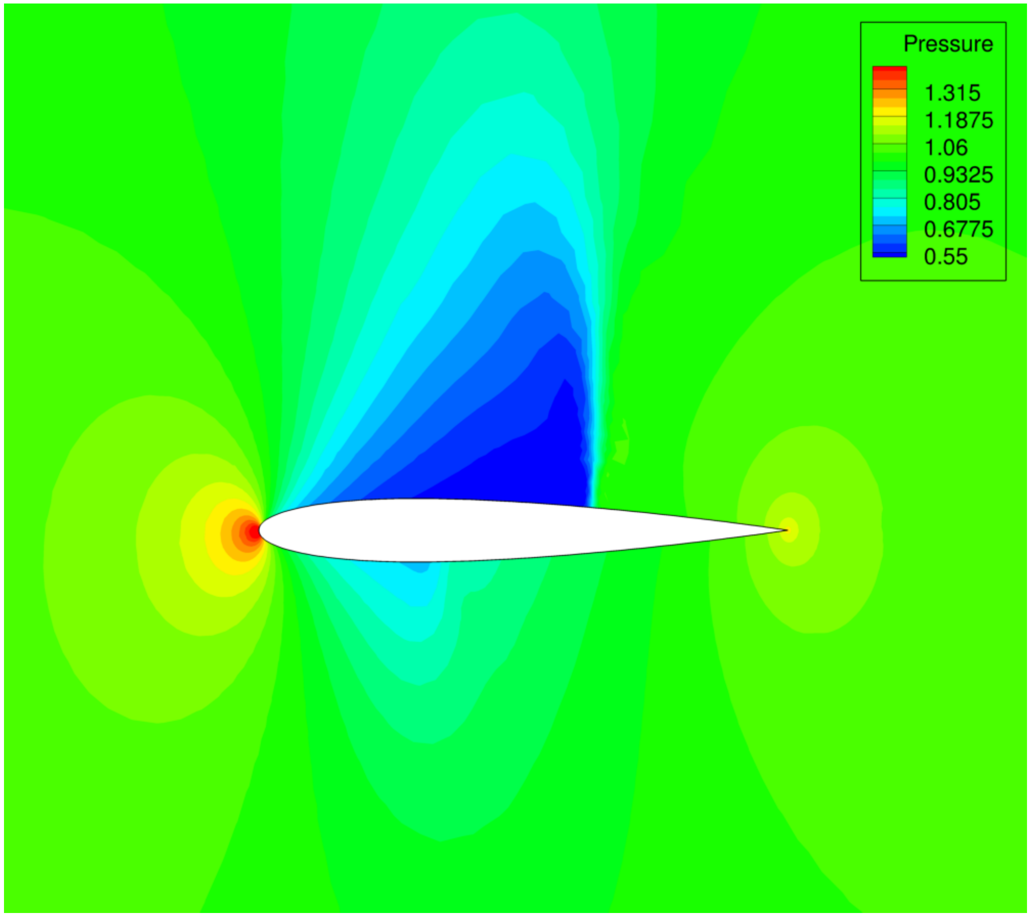
# An example where it doesn't work

$$f(x_1, x_2) = 5x_1 + \sin(10\pi x_2)$$

$$\boldsymbol{C} = \begin{bmatrix} 25 & 0 \\ 0 & 526 \end{bmatrix}$$



# An example where it works




**DRAG COEFFICIENT**

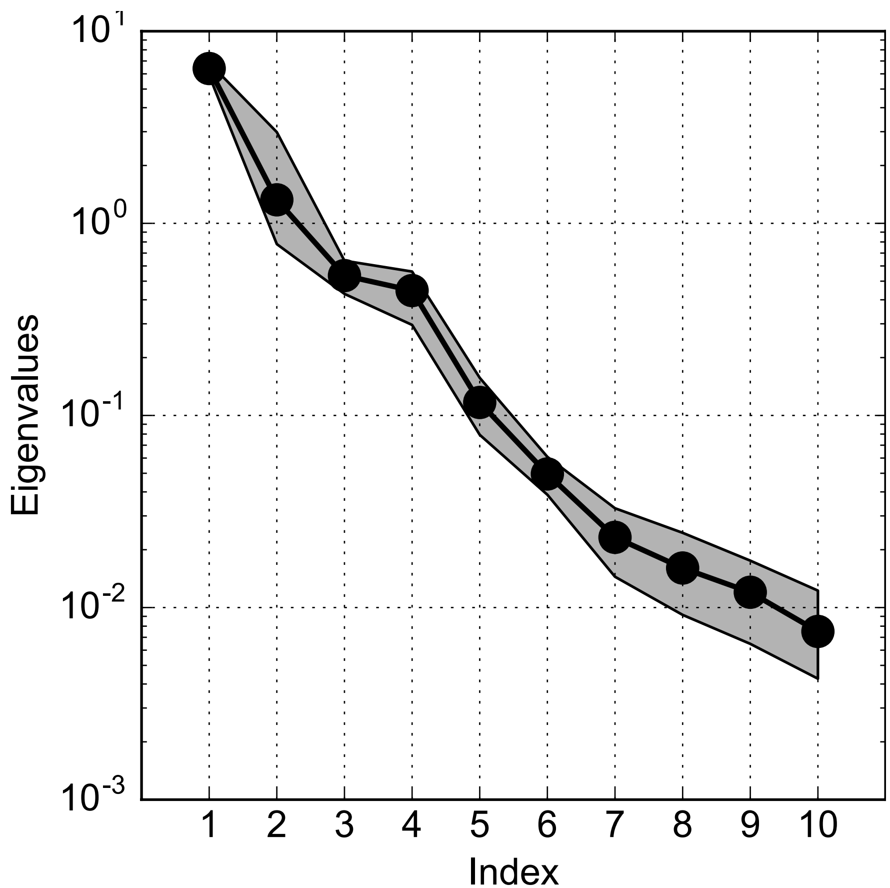
as a function of

**18** shape parameters

Uniform on a hypercube

SU2 CFD solver with adjoint solver for gradients

# An example where it works




**DRAG COEFFICIENT**

as a function of

**18** shape parameters

Uniform on a hypercube

SU2 CFD solver with adjoint solver for gradients

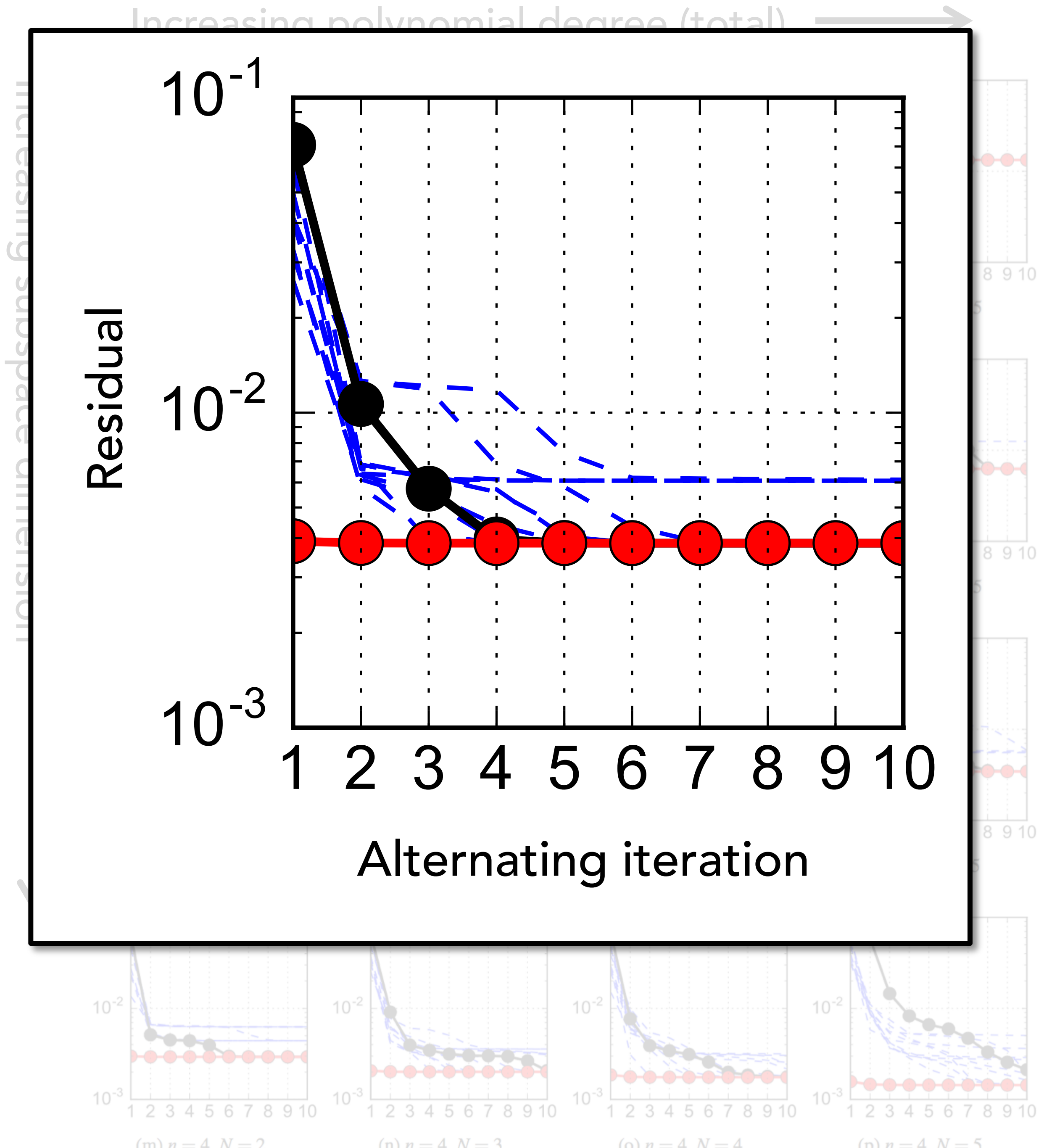
Recall: $$\boldsymbol{C} = \int \nabla f(\mathbf{x})\, \nabla f(\mathbf{x})^T \rho(\mathbf{x})\, d\mathbf{x} = \boldsymbol{W} \Lambda \boldsymbol{W}^T$$

Residual as a function of alternating iteration for different starting subspaces

**RANDOM**

**IDENTITY**

**ACTIVE SUBSPACE**

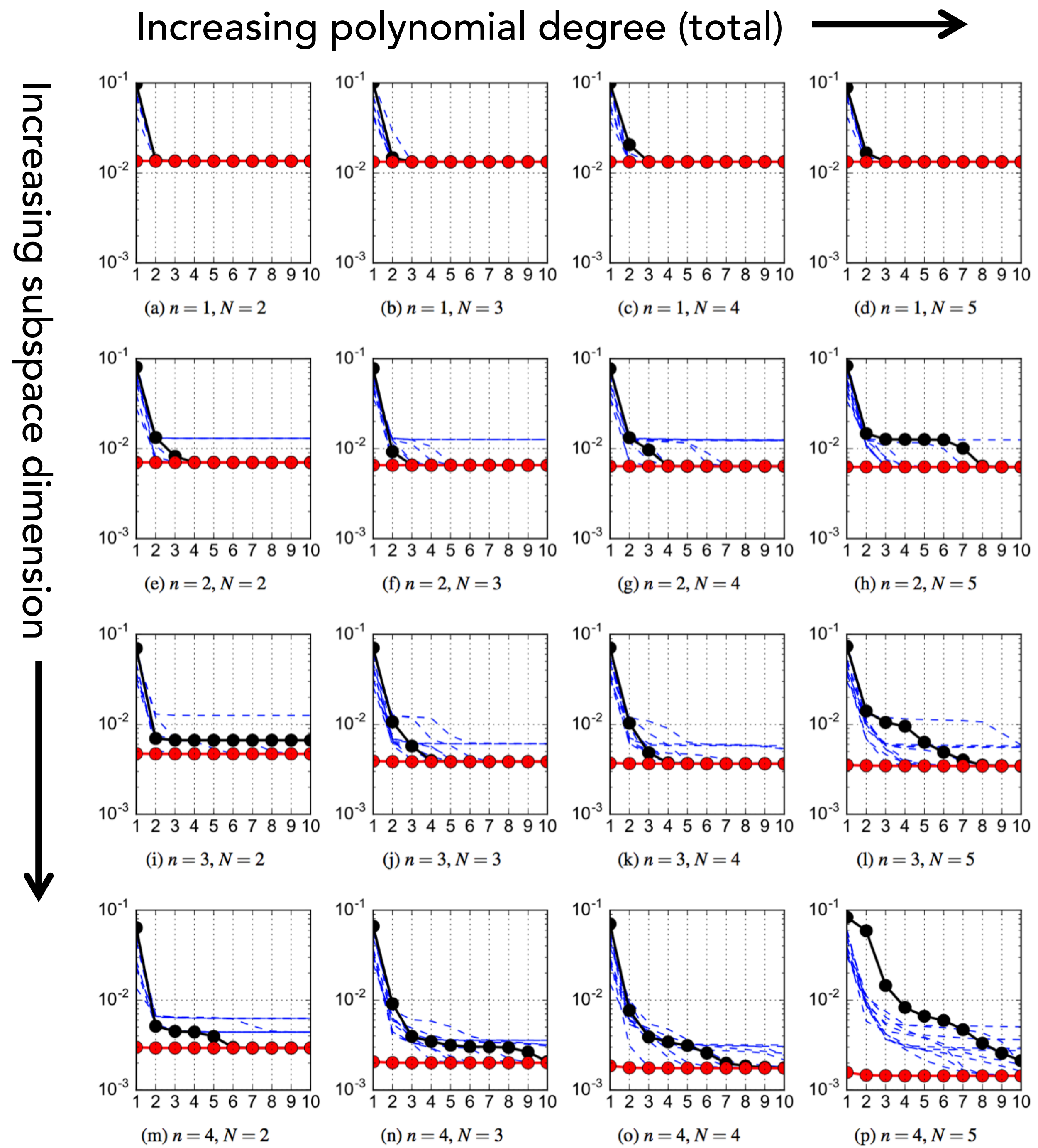


# Residual as a function of alternating iteration for different starting subspaces

**RANDOM**

**IDENTITY**

**ACTIVE SUBSPACE**



# TAKE-HOMES

An active subspace is a type of low-dimensional structure in a function of several variables.

We have tools for identifying and exploiting active subspaces for parameter studies.

Active subspaces appear in a wide range of physical models.

Active subspaces are closely related to ridge approximation.

# QUESTIONS?

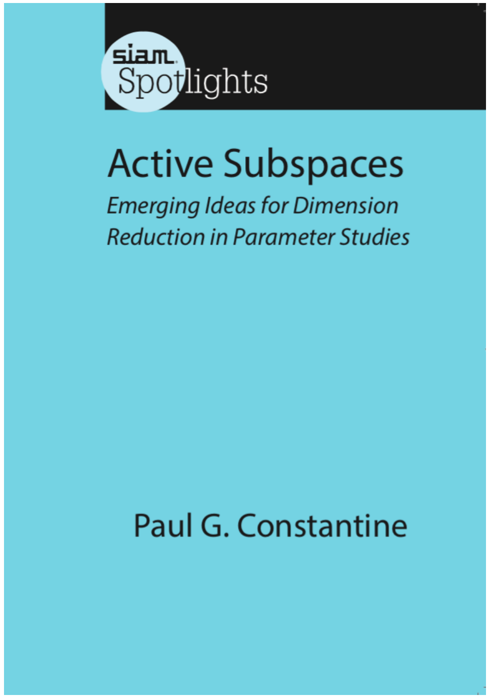
How do active subspaces relate to [insert method]?

How do I compute active subspaces?

What if I don't have gradients?

What kinds of models does this work on?

Active Subspaces
SIAM (2015)




**PAUL CONSTANTINE**
Ben L. Fryrear Assistant Professor
Colorado School of Mines
activesubspaces.org
@DrPaulynomial

# BACK UP SLIDES

## Sufficient Dimension Reduction in regression

Assume: $y = f(\mathbf{x}) + \epsilon$

Assume: $\mathbb{P}(y|\mathbf{x}) = \mathbb{P}(y|\boldsymbol{A}^T\mathbf{x})$

Given $(y_i, \mathbf{x}_i)$, find $\boldsymbol{A}$

## Projection Pursuit Regression, Neural Nets, Ridge functions

$$\underset{\boldsymbol{A},\theta}{\text{minimize}} \int (f(\mathbf{x}) - g(\boldsymbol{A}^T\mathbf{x}, \theta))^2 \, \rho \, d\mathbf{x}$$

## Fisher Information Theory

$$\int \nabla_\theta^2 \log \mathcal{L}(\mathbf{x}, \theta) \, \rho(\mathbf{x}) \, d\mathbf{x}$$

## Principal Components / Regression, Karhunen-Loéve

$$\int \mathbf{x} \, \mathbf{x}^T \rho(\mathbf{x}) \, d\mathbf{x}$$

# References and related work


**"Sufficient dimension reduction"**

- Cook. *Regression Graphics.* (1998/2009)
- K.C. Li. *Sliced inverse regression for dimension reduction.* (1991)
- K.C. Li. *On principal Hessian directions for data visualization and dimension reduction* (1992)

**In approximation theory**

- Fornassier, Schass, Vybiral. *Learning functions of few arbitrary linear parameters in high dimensions.* FOCM (2012)
- Mayer, Ullrich, Vybiral. *Entropy and sampling numbers of classes of ridge functions.* Constructive Approximation (2014)

**In UQ**

- Tipireddy, Ghanem. *Basis adaptation in homogeneous chaos spaces.* JCP (2014)
- Stoyanov, Webster. *A gradient-based sampling approach for dimension reduction of partial differential equations with stochastic coefficients.* IJ4UQ (2015)
- Lei, Yang, Lei, Zheng, Lin, Baker. *Constructing Surrogate Models of Complex Systems with Enhanced Sparsity: Quantifying the Influence of Conformational Uncertainty in Biomolecular Solvation.* SIAM MMS (2015)

**In engineering applications**

- Abdel-Khalik, Bang, Wang. *Overview of hybrid subspace methods for uncertainty quantification, sensitivity analysis.* Annals of Nuclear Energy (2013)
- Berguin, Rancourt, Mavris. *A method for high-dimensional design space exploration of expensive functions with access to gradient information.* AIAA-2014-2174
- Russi. *Uncertainty Quantification with Experimental Data and Complex System Models,* Ph.D. thesis (2010)


**In statistics/machine learning**: "principal pursuit regression," "neural nets"

# Discover the active subspace with random sampling.

Draw samples: $\mathbf{x}_j \sim \rho$

Compute: $f_j = f(\mathbf{x}_j)$ and $\nabla f_j = \nabla f(\mathbf{x}_j)$

Approximate with Monte Carlo

$$\boldsymbol{C} \approx \frac{1}{N} \sum_{j=1}^{N} \nabla f_j \, \nabla f_j^T = \hat{\boldsymbol{W}} \hat{\Lambda} \hat{\boldsymbol{W}}^T$$

Equivalent to SVD of samples of the gradient

$$\frac{1}{\sqrt{N}} \begin{bmatrix} \nabla f_1 & \cdots & \nabla f_N \end{bmatrix} = \hat{\boldsymbol{W}} \sqrt{\hat{\Lambda}} \hat{\boldsymbol{V}}^T$$

Called an *active subspace method* in T. Russi's 2010 Ph.D. thesis, *Uncertainty Quantification with Experimental Data in Complex System Models*

## How many gradient samples?

Bound on gradient norm squared

Dimension

$$N = \Omega\left( \frac{L^2 \lambda_1}{\lambda_k^2 \varepsilon^2} \log(m) \right) \implies |\lambda_k - \hat{\lambda}_k| \leq \varepsilon \lambda_k$$

Relative accuracy

(with high probability)

Using Gittens and Tropp (2011)

## How many gradient samples?

Bound on gradient norm squared

Dimension

(with high probability)

$$N = \Omega\left(\frac{L^2}{\lambda_1 \varepsilon^2}\log(m)\right) \quad \Longrightarrow \quad \mathrm{dist}\left(\boldsymbol{W}_1, \hat{\boldsymbol{W}}_1\right) \leq \frac{4\lambda_1\varepsilon}{\lambda_n - \lambda_{n+1}}$$

Relative accuracy

Spectral gap

Gittens and Tropp (2011), Golub and Van Loan (1996), Stewart (1973)

# Let's be abundantly clear about the problem we are trying to solve.

Low-rank approximation of the collection of gradients:

$$\frac{1}{\sqrt{N}}\begin{bmatrix}\nabla f_1 & \cdots & \nabla f_N\end{bmatrix} \approx \hat{\boldsymbol{W}}_1\sqrt{\hat{\Lambda}_1}\hat{\boldsymbol{V}}_1^T$$

Low-dimensional linear approximation of the gradient:

$$\nabla f(\mathbf{x}) \approx \hat{\boldsymbol{W}}_1\,\mathbf{a}(\mathbf{x})$$

**Approximate** a function of many variables by a function of a few linear combinations of the variables:

$$f(\mathbf{x}) \approx g\left(\hat{\boldsymbol{W}}_1^T\mathbf{x}\right)$$