Exploiting active subspaces for optimization of physics-based models

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

activesubspaces.org
@DrPaulynomial

SLIDES AVAILABLE UPON REQUEST

DISCLAIMER: These slides are meant to complement the oral presentation. Use out of context at your own risk.
minimize \( f(x) \) \( \rightarrow \) design objective
subject to \( x \in [-1, 1]^m \) \( \rightarrow \) model parameters

**PROPERTIES**
- Numerical approximation of PDE “under the hood”
- PDE models a complex physical system
- Numerical “noise”
- Typically no gradients or Hessians
- Expensive to evaluate (minutes-to-months)
- More “black box” than PDE-constrained optimization

**APPLICATIONS** I’ve worked on
- Design of aerospace systems
- Hydrologic system modeling
- Renewable energy systems
\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad x \in [-1, 1]^m
\end{align*}
\]

\textbf{INTRACTABLE}, in general!
- Requires dense “trial points” (Theorem 1.3, Törn and Žilinskas (1987))
- Curse of dimensionality (Traub and Werschulz (1998))

\textbf{VAST LITERATURE} on response surface or model-based approaches, e.g.,
- Jones, Schonlau, and Welch (1998)
- Jones [“taxonomy”] (2001)
- Shan and Wang [“HEB” review] (2010)

And many, many other heuristics…
"The greatest value of a picture is when it forces us to notice what we never expected to see."

"Even more understanding is lost if we consider each thing we can do to data only in terms of some set of very restrictive assumptions under which that thing is best possible---assumptions we know we CANNOT check in practice."

"Exploratory data analysis is detective work work work ..."
Quantifying safety margins in a multiphysics scramjet model

seven parameters characterizing the operating conditions

What is the range of pressures at the channel exit?

- 9500 CPU hours per run
- no gradients or Hessians
- noisy function evaluations

Constantine, Emory, Larsson, and Iaccarino (2015)
Quantifying safety margins in a multiphysics scramjet model

\[ \mathbf{x}_{\text{min}} = \left\{ \begin{array}{l} \arg\min_{\mathbf{x}} \mathbf{w}^T \mathbf{x} \\ \text{subject to } \mathbf{x} \in [-1, 1]^m \end{array} \right\} = -\text{sign}(\mathbf{w}), \quad f_{\text{min}} = f(\mathbf{x}_{\text{min}}) \]

Constantine, Emory, Larsson, and Iaccarino (2015)
Design a jet nozzle under uncertainty
(DARPA SEQUOIA project)

youtu.be/Fek2HstkJFVc

Alonso, Eldred, Constantine et al. (2017)
IDEA  Given \( w \), plot \( f(x) \) versus \( w^T x \)

QUESTION  How to get \( w \) ?
### How to get $w$?

<table>
<thead>
<tr>
<th>Method</th>
<th>Formula</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gradient of least-squares linear model</td>
<td>$f(x) \approx a + b^T x$</td>
<td>[Li and Duan (1989)]</td>
</tr>
<tr>
<td></td>
<td>$w = b / |b|$</td>
<td>[Li (1991)]</td>
</tr>
<tr>
<td>Sliced Inverse Regression (SIR)</td>
<td>$\text{Sliced Inverse Regression (SIR): first eigenvector of}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\text{Cov}[\mathbb{E}[x</td>
<td>f]]$</td>
</tr>
<tr>
<td>Sliced Average Variance Estimation (SAVE)</td>
<td>$\mathbb{E}[ (I - \text{Cov}[x</td>
<td>f])^2 ]$</td>
</tr>
<tr>
<td>Projection Pursuit Regression (PPR), Ridge Approximation</td>
<td>$\text{Projection Pursuit Regression (PPR), Ridge Approximation}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\min_{w,g} | f(x) - g(w^T x) |$</td>
<td>[Friedman and Stuetzle (1981), Constantine et al. (2016)]</td>
</tr>
<tr>
<td>Active Subspaces: first eigenvector of</td>
<td>$\mathbb{E}[ \nabla f(x) \nabla f(x)^T ]$</td>
<td></td>
</tr>
<tr>
<td>Principal Hessian Directions (pHd): first eigenvector of</td>
<td>$\mathbb{E}[ \nabla^2 f(x) ]$</td>
<td>[Hristache et al. (2001), Constantine et al. (2014)]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[Li (1992)]</td>
</tr>
</tbody>
</table>
How to get $\mathbf{w}$?

Gradient of least-squares linear model

$$f(\mathbf{x}) \approx a + \mathbf{b}^T \mathbf{x}$$

$$\mathbf{w} = \mathbf{b} / \|\mathbf{b}\|$$

[Li and Duan (1989)]

NOTES

+ Maximizes squared correlation
+ Cheap to fit
- Misses quadratic-like behavior

Sliced Inverse Regression (SIR): first eigenvector of

$$\text{Cov}[\mathbb{E}[\mathbf{x}|f]]$$

Sliced Average Variance Estimation (SAVE): first eigenvector of

$$\mathbb{E}[(\mathbf{I} - \text{Cov}[\mathbf{x}|f])^2]$$

[Cook and Weisberg (1991)]

Projection Pursuit Regression (PPR), Ridge Approximation

$$\min_{\mathbf{w},g} \|f(\mathbf{x}) - g(\mathbf{w}^T \mathbf{x})\|$$

[Li (1992)]

Principal Hessian Directions (pHd): first eigenvector of

$$\mathbb{E}[\nabla^2 f(\mathbf{x})]$$

[Friedman and Stuetzle (1981), Constantine et al. (2016)]

[Hristache et al. (2001), Constantine et al. (2014)]
**How to get \( \mathbf{w} \)?**

Gradient of least-squares linear model

\[
f(x) \approx a + b^T x
\]

\[
\mathbf{w} = \frac{\mathbf{b}}{\|\mathbf{b}\|}
\]

[Li and Duan (1989)]

<table>
<thead>
<tr>
<th>Sliced Inverse Regression (SIR): first eigenvector of</th>
<th>Sliced Average Variance Estimation (SAVE): first eigenvector of</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Cov} [\mathbb{E} [\mathbf{x}</td>
<td>f ]] )</td>
</tr>
</tbody>
</table>

[Li (1991)]

Projection Pursuit Regression (PPR), Ridge Approximation

\[
\min_{w, g} \| f(x) - g(w^T x) \|
\]

[Friedman and Stuetzle (1981), Constantine et al. (2016)]

**NOTES**

Methods for *inverse regression*
## How to get $w$?

| Gradient of least-squares linear model | Sliced Inverse Regression (SIR): first eigenvector of $f(x) \approx a + b^T x$ | Sliced Average Variance Estimation (SAVE): first eigenvector of $E\left[ (I - \text{Cov}[x|f])^2 \right]$ |
|----------------------------------------|---------------------------------------------------------------------------------|----------------------------------------------------------------------------------|
| $w = b / \|b\|$                       | $\text{Cov}[E[x|f]]$                                                           | $[\text{Li (1991)}]$                                                             |
| [Li and Duan (1989)]                  | [Li (1991)]                                                                    | [Cook and Weisberg (1991)]                                                       |

**NOTES**

- Known as *sufficient dimension reduction*

Principal Hessian Directions (pHd): first eigenvector of $E[\nabla^2 f(x)]$

[Li (1992)]
How to get \( w \)?

**NOTES**

- Two of four *average derivative functionals* from Samarov (1993)
  - Require derivatives
- Use model-based derivative approximations, if derivatives not available

Active Subspaces: first eigenvector of

\[
\mathbb{E}\left[ \nabla f(x) \nabla f(x)^T \right]
\]

Principal Hessian Directions (pHd): first eigenvector of

\[
\mathbb{E}\left[ \nabla^2 f(x) \right]
\]

[Hristache et al. (2001), Constantine et al. (2014)]

[Li (1992)]
How to get $w$?

Gradient of least-squares linear model

$g(w) = \| f(x) - g(w^T x) \|^2$

NOTES

Not all eigenvector-based techniques are PCA!

[Sliced Inverse Regression (SIR): first eigenvector of $\text{Cov} [\mathbb{E}[ x | f ]]$]

[Sliced Average Variance Estimation (SAVE): first eigenvector of $\mathbb{E} [ (I - \text{Cov} [ x | f ])^2 ]$]

[Li (1991)]

[Sliced Average Variance Estimation (SAVE): first eigenvector of $\mathbb{E} [ (I - \text{Cov} [ x | f ])^2 ]$]

[Cook and Weisberg (1991)]

[Active Subspaces: first eigenvector of $\mathbb{E} [ \nabla f(x) \nabla f(x)^T ]$]

[Principal Hessian Directions (pHd): first eigenvector of $\mathbb{E} [ \nabla^2 f(x) ]$]

[Hristache et al. (2001), Constantine et al. (2014)]

[Li (1992)]

[Projection Pursuit Regression (PPR), Ridge Approximation [Friedman and Stuetzle (1981), Constantine et al. (2016)]]

$\min_{w,g} \| f(x) - g(w^T x) \|^2$
How to get $w$?

Gradient of least-squares linear model
\[ f(x) \approx a + b^T x \]
\[ w = b / \|b\| \]
[Li and Duan (1989)]

Sliced Inverse Regression (SIR): first eigenvector of
\[ \text{Cov}[E[x|f]] \]
[Li (1991)]

Sliced Average Variance Estimation (SAVE): first eigenvector of
\[ E[(I - \text{Cov}[x|f])^2] \]
[Cook and Weisberg (1991)]

Principal Hessian Directions (pHd): first eigenvector of
\[ E[r^2f(x)r^Tf(x)] \]
[Li (1992)]

Active Subspaces: first eigenvector of
\[ E[rf(x)rf(x)^T] \]
[Hristache et al. (2001), Constantine et al. (2014)]

Projection Pursuit Regression (PPR), Ridge Approximation
\[ \min_{w, g} \| f(x) - g(w^T x) \| \]
[Friedman and Stuetzle (1981), Constantine et al. (2016)]

NOTES

- Related to neural nets; see Hastie et al. ESL (2009)
- Different from ridge recovery; see Fornasier et al. (2012)
### How to get \( \mathbf{w} \)?

| Gradient of least-squares linear model | Sliced Inverse Regression (SIR): first eigenvector of \( \text{Cov}[\mathbb{E}[\mathbf{x}|f]] \) | Sliced Average Variance Estimation (SAVE): first eigenvector of \( \mathbb{E}[(I - \text{Cov}[\mathbf{x}|f])^2] \) |
|---------------------------------------|---------------------------------------------------------------------------------|--------------------------------------------------------------------------------|
| \( f(\mathbf{x}) \approx a + b^T \mathbf{x} \) | \( \mathbf{w} = b / \|b\| \) | \([\text{Li (1991)}]\) |
| \([\text{Li and Duan (1989)}]\) | \([\text{Li (1991)}]\) | \([\text{Cook and Weisberg (1991)}]\) |

<table>
<thead>
<tr>
<th>Projection Pursuit Regression (PPR), Ridge Approximation</th>
<th>Active Subspaces: first eigenvector of ( \mathbb{E}[\nabla f(\mathbf{x}) \nabla f(\mathbf{x})^T] )</th>
<th>Principal Hessian Directions (pHd): first eigenvector of ( \mathbb{E}[\nabla^2 f(\mathbf{x})] )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \min_{\mathbf{w},g} | f(\mathbf{x}) - g(\mathbf{w}^T \mathbf{x}) | )</td>
<td>([\text{Friedman and Stuetzle (1981), Constantine et al. (2016)}])</td>
<td>([\text{Hristache et al. (2001), Constantine et al. (2014)}])</td>
</tr>
<tr>
<td>([\text{Friedman and Stuetzle (1981), Constantine et al. (2016)}])</td>
<td>([\text{Hristache et al. (2001), Constantine et al. (2014)}])</td>
<td>([\text{Li (1992)}])</td>
</tr>
</tbody>
</table>

**NOTES** Regression or approximation?

Glaws, Constantine, and Cook (2017)
So you found w ...

**QUESTION**  Is it any good?
Why isn’t it perfectly univariate?

The function varies along directions orthogonal to the computed $\mathbf{w}$

The function varies due to other variables (“noise”)

The computed $\mathbf{w}$ is wrong because you used the wrong method

The computed $\mathbf{w}$ is a poor numerical estimate
Why isn’t it perfectly univariate?

The function varies along directions orthogonal to the computed \( \mathbf{w} \)

\[
\mathbb{E}[\nabla f(x) \nabla f(x)^T] = \mathbf{W} \Lambda \mathbf{W}^T
\]

The function varies due to other variables (“noise”)

The computed \( \mathbf{w} \) is wrong because you used the wrong method

The computed \( \mathbf{w} \) is a poor numerical estimate

NOTES

Check with

- eigenvalues, e.g.,

\[
\mathbb{E}[\nabla f(x) \nabla f(x)^T] = \mathbf{W} \Lambda \mathbf{W}^T
\]

- additional function evaluations (expensive)
Why isn’t it perfectly univariate?

The function varies along directions orthogonal to the computed \( w \).

The function varies due to other variables ("noise").

The computed \( w \) is wrong because you used the wrong method.

The computed \( w \) is a poor numerical estimate.

---

**NOTES**

Check for computational "noise"; see Moré and Wild (2011)
Why isn’t it perfectly univariate?

The function varies along directions orthogonal to the computed $w$.

The function varies due to other variables ("noise").

The computed $w$ is wrong because you used the wrong method.

The computed $w$ is a poor numerical estimate.

**NOTES**

Try multiple approaches for computing $w$, if possible.
Why isn’t it perfectly univariate?

The function varies along directions orthogonal to the computed $w$

The function varies due to other variables ("noise")

The computed $w$ is wrong because you used the wrong method

The computed $w$ is a poor numerical estimate

**NOTES**

- Take more samples; e.g., see Constantine and Gleich (2015)

\[
\mathbb{E} [\nabla f(x) \nabla f(x)^T ] \\
\approx \frac{1}{N} \sum_{i=1}^{N} \nabla f(x_i) \nabla f(x_i)^T
\]

- Account for “noise”
If the objective function is (strictly) monotonic, then replace the “black box” objective with a linear function.

\[
\arg\min_{x} w^T x \quad \text{subject to} \quad x \in [-1, 1]^m \implies -\text{sign}(w)
\]

Testing monotonicity of regression
- Bowman et al. (1998)
- Ghoshal et al. (2000)

Can we automatically test for monotonicity?

Regression or approximation?
Use a bootstrap (sampling with replacement) to assess sensitivity of weights with respect to the "data."

**NOTES**

Weight close to zero indicates
- high sensitivity in *minimizer*
- low sensitivity in *minimum*

Weight can give sensitivity information; see Constantine and Diaz (2017)
Do real world models have such structure?
Evidence of structure: 
Multiphysics hypersonic scramjet model

Constantine, Emory, Larsson, and Iaccarino (2015)
Evidence of structure: Integrated hydrologic model

Jefferson, Gilbert, Constantine, and Maxwell (2015); Jefferson, Constantine, and Maxwell (2017)
Evidence of structure:
Transonic wing design

Wing perturbations

Lukaczyk, Constantine, Palacios, and Alonso (2014); Constantine (2015)
Evidence of structure:
In-host HIV dynamics

Loudon and Pankavich (2017)
Evidence of structure: Solar-cell circuit model

Constantine, Zaharatos, and Campanelli (2015)
Evidence of structure:
Atmospheric re-entry vehicle

Cortesi, Constantine, Magin, and Congedo (In prep.)
Evidence of structure:
Magnetohydrodynamics generator model

Glaws, Constantine, Shadid, and Wildey (2017)
Evidence of structure: Lithium-ion battery model

Constantine and Doostan (2017)
Evidence of structure: Automobile design
Evidence of no 1-d structure:
A subsurface hydrology problem

Domain

Hydraulic conductivities

Hortonian Case

Cumulative Runoff Volume [cubic meters]

Active Variable

Gilbert, Jefferson, Constantine, and Maxwell (2016)
### Active Subspace Data Sets — Edit

<table>
<thead>
<tr>
<th>Branch: master</th>
<th>New pull request</th>
<th>Create new file</th>
<th>Upload files</th>
<th>Find file</th>
<th>Clone or download</th>
</tr>
</thead>
</table>

- **paulcon committed on GitHub** Merge pull request **#28** from ryan-kelley-howard/master

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Description</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atacamac</td>
<td>Title, author, and reference changes</td>
<td>11 days ago</td>
</tr>
<tr>
<td>Ebola</td>
<td>Title, author, and reference changes</td>
<td>11 days ago</td>
</tr>
<tr>
<td>HIV</td>
<td>Title, author, and reference changes</td>
<td>11 days ago</td>
</tr>
<tr>
<td>HyShotll</td>
<td>Title, author, and reference changes</td>
<td>11 days ago</td>
</tr>
<tr>
<td>Hydrology</td>
<td>Title, author, and reference changes</td>
<td>11 days ago</td>
</tr>
</tbody>
</table>

Latest commit **da61ac5 10 days ago**
WHY IS THIS HAPPENING???

Here’s my hunch …

(1) monotonicity

(2) small range of parameters
SOME EXTENSIONS

Two-dimensional scatter plots

Ridge functions and ridge approximations

\[ f(x) \approx g(U^T x) \]

\[ U^T : \mathbb{R}^m \rightarrow \mathbb{R}^n \]

\[ g : \mathbb{R}^n \rightarrow \mathbb{R} \]

See, e.g., Wang et al., Bayesian Optimization in a Billion Dimensions via Random Embeddings (JAIR, 2016)
SIAM Workshop on Parameter Space Dimension Reduction (DR17)

July 9-10, 2017
Omni William Penn Hotel
Pittsburgh, Pennsylvania USA

www.siam.org/meetings/dr17/
TAKE HOME

Check your optimization problem for exploitable (low-d, monotonic) structure with exploratory, graphical analysis!

QUESTIONS?
Ask me about the elliptic PDE problem!

PAUL CONSTANTINE
Ben L. Fryrear Assistant Professor
Colorado School of Mines
activesubspaces.org
@DrPaulynomial
BACK UP SLIDES
The parameterized PDE

\[-\nabla \cdot (a \nabla u) = 1, \ s \in \mathcal{D}\]
\[u = 0, \ s \in \Gamma_1\]
\[-n \cdot a \nabla u = 0, \ s \in \Gamma_2\]

The coefficients

\[\log(a(s, x)) = \sum_{k=1}^{m} \sqrt{\theta_k} \phi_k(s) x_k\]

\[\text{Cov}(s_1, s_2) = \sigma^2 \exp \left( \frac{-\|s_1 - s_2\|_1}{2 \ell} \right)\]

The quantity of interest

\[f(x) = \int u(s, x) \, d\Gamma_2\]
**Long length scale, \( \ell = 1 \)**

\[
\log(a(s, x))
\]

\[
u(s, x)
\]
Short length scale, $\ell = 0.01$

$$\log(a(s, x))$$

$$u(s, x)$$
| \( f(x) \) | PDE solution’s spatial average along the Neumann boundary |
| \( \rho(x) \) | Standard Gaussian density |
| \( \nabla f(x) \) | Gradient computed with adjoint equations |
Plotting $f(x_j)$ versus $\hat{w}_1^T x_j$

Long length scale, $\ell = 1$

Short length scale, $\ell = 0.01$

Remember the goal!

$$f(x) \approx g(\hat{w}_1^T x)$$
Eigenvalues of $\frac{1}{N} \sum_{j=1}^{N} \nabla f_j \nabla f_j^T$

Long length scale, $\ell = 1$

Short length scale, $\ell = 0.01$
Estimates of subspace error $\varepsilon = \text{dist}(W_1, \hat{W}_1)$

Long length scale, $\ell = 1$

Short length scale, $\ell = 0.01$
First eigenvector of \( \frac{1}{N} \sum_{j=1}^{N} \nabla f_j \nabla f_j^T \)

Long length scale, \( \ell = 1 \)

Short length scale, \( \ell = 0.01 \)