Randomized Computation of Active Subspaces

Ilse Ipsen

Joint with: John T. Holodnak & Ralph C. Smith

North Carolina State University
Raleigh, NC, USA

Research supported by DARPA XData and NSF CCF

Motivation

Given: Differentiable function $f: \mathbb{R}^m \to \mathbb{R}$ where m large

Want: Influential parameters of f

1. Detect active subspace $\mathcal{S} \subset \mathbb{R}^m$ where f most sensitive to change (varies strongly)
2. Approximate f by response surface over $\mathcal{S}$

Existing Work:

Active subspaces [Russi 2010]
Stochastic PDEs [Constantine et al. 2012, 2014], [Stoyanov et al. 2014]
Reduced-order nonlinear models [Bang et al. 2012]
Airfoil design and manufacturing [Namura et al. 2015], [Chen et al. 2011]
Combustion [Bauernheim et al. 2014], [Constantine et al. 2011]
Solar cells [Constantine et al. 2014]

Idea

Given: Function $f: \mathbb{R}^m \to \mathbb{R}$

1. From $\nabla f(x)$ construct "sensitivity" matrix $E \in \mathbb{R}^{m \times m}$
2. Dominant eigenvectors of $E \Rightarrow$ active subspace $\mathcal{S}$

Problem: Elements of E too expensive to compute

(high-dimensional integrals)

4. Approximate E by Monte Carlo: $\widehat{E} \in \mathbb{R}^{m \times m}$
5. Dominant eigenvectors of $\widehat{E} \Rightarrow$ approximate subspace $\widehat{\mathcal{S}}$

Our contribution: Probabilistic bound for $\sin \angle(\mathcal{S}, \widehat{\mathcal{S}})$

Tight if: E has low numerical rank and large eigenvalue gap

Overview

1. Assumptions
   - "Sensitivity" matrix E
   - Active subspace $\mathcal{S}$
   - Monte Carlo approximation $\widehat{E}$
   - Approximate subspace $\widehat{\mathcal{S}}$
2. Accuracy of $\widehat{\mathcal{S}}$
   - Structural (deterministic) bound for subspace angle
   - Matrix concentration inequality
   - Probabilistic bound for number of Monte Carlo samples

Assumptions

The function is somewhat nice

- $f : \mathbb{R}^m \to \mathbb{R}$ continuously differentiable
- Lipschitz constant $\|\nabla f(x)\| \leq L$ (2 norm)

Monte Carlo sampling

- Random vectors $\mathbf{X} \in \mathbb{R}^m$ with probability density $\rho(x)$
- Expected value of function h with respect to $\mathbf{X}$

$$\mathbb{E}[h(\mathbf{X})] \equiv \int_{\mathbb{R}^m} h(x)\,\rho(x)dx$$

"Sensitivity" Matrix E

Informative directional derivatives

$$E \;\equiv\; \int_{\mathbb{R}^m} \nabla f(x)(\nabla f(x))^T \, \rho(x) dx$$

- $E \in \mathbb{R}^{m\times m}$ symmetric positive semi-definite
- Eigenvalue decomposition $E = V \Lambda V^T$
- Eigenvectors $V = \begin{pmatrix} v_1 & \dots & v_m \end{pmatrix}$

  v_j is direction of sensitivity of f
- Eigenvalues $\Lambda = \mathrm{diag}\begin{pmatrix} \lambda_1 & \cdots & \lambda_m \end{pmatrix}$

  $\lambda_j \;=\; \mathbb{E}\left[(v_j^T \nabla f(\mathbf{X}))^2\right]$ average sensitivity along v_j

Active Subspace $\mathcal{S}$

Dominant eigenvalues of $E = V \Lambda V^T$

$$\Lambda = \operatorname{diag}\begin{pmatrix} \lambda_1 & \cdots & \lambda_k & \lambda_{k+1} & \cdots & \lambda_m \end{pmatrix}$$

- Large eigenvalue gap

$$\lambda_1 \geq \cdots \geq \lambda_k \gg \lambda_{k+1} \geq \cdots \geq \lambda_m$$

- k dominant eigenvalues λ_j: Indicators of high sensitivity
- k dominant eigenvectors v_j: Directions of high sensitivity

Orthonormal basis for active subspace

$$\mathcal{S} \equiv \operatorname{range}\begin{pmatrix} v_1 & \cdots & v_k \end{pmatrix}$$

Monte Carlo Approximation $\widehat{E}$

- Sample $n \ll m$ training points $x_j \in \mathbb{R}^m$ according to $\rho(x)$

$$\widehat{E} = \frac{1}{n} \sum_{j=1}^{n} \nabla f(x_j) \left(\nabla f(x_j)\right)^T$$

- Eigenvalue decomposition $\widehat{E} = \widehat{V} \widehat{\Lambda} \widehat{V}^T$

$$\widehat{\Lambda} = \mathrm{diag}\begin{pmatrix} \widehat{\lambda}_1 & \cdots & \widehat{\lambda}_k & \widehat{\lambda}_{k+1} & \cdots & \widehat{\lambda}_m \end{pmatrix}$$

- Assume: Eigenvalue gap in same location as for E

$$\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_k \gg \widehat{\lambda}_{k+1} \geq \cdots \geq \widehat{\lambda}_m$$

Orthonormal basis for approximate subspace

$$\widehat{\mathcal{S}} \equiv \mathrm{range}\begin{pmatrix} \widehat{v}_1 & \cdots & \widehat{v}_k \end{pmatrix}$$

Accuracy of Approximate Subspace

Approach

1. Structural (deterministic) bound

   Bound $\sin\angle(\mathcal{S}, \widehat{\mathcal{S}})$ in terms of $\|\widehat{E} - E\|$

2. Probabilistic bound

   Bound $\|\widehat{E} - E\|$ in terms of sampling amount n

3. Combine the two bounds

   Sampling amount n so that $\sin\angle(\mathcal{S}, \widehat{\mathcal{S}}) \leq \epsilon$

Structural Bound: Subspace Perturbation

based on [Stewart 1973]

- Eigenvalues of E: $\lambda_1 \geq \cdots \geq \lambda_k > \lambda_{k+1} \geq \cdots \geq \lambda_m$
- Active subspace: $\mathcal{S} = \text{range}\begin{pmatrix} v_1 & \cdots & v_k \end{pmatrix}$
- Approximate subspace: $\widehat{\mathcal{S}} = \text{range}\begin{pmatrix} \widehat{v}_1 & \cdots & \widehat{v}_k \end{pmatrix}$
- Small enough perturbation: $\|\widehat{E} - E\| \leq \frac{1}{4}(\lambda_k - \lambda_{k+1})$

Then

$$\sin\angle(\mathcal{S}, \widehat{\mathcal{S}}) \;\leq\; 4\,\frac{\|\widehat{E} - E\|}{\lambda_k - \lambda_{k+1}}$$

If $\lambda_{k+1} - \lambda_k \gg 0$ then active subspace $\mathcal{S}$ well-conditioned

Probabilistic Bound: Matrix Perturbation

- Want: Probabilistic bound for $\|\widehat{E} - E\|$
- Exact: $E \;=\; \int_{\mathbb{R}^m} \nabla f(x)(\nabla f(x))^T \, \rho(x) dx$
- Monte Carlo approximation: $\widehat{E} = \frac{1}{n} \sum_{j=1}^{n} \nabla f(x_j) \, (\nabla f(x_j))^T$
- Idea: $\widehat{E}$ is average of matrix-valued random variables

$$\nabla f(x_j) \, (\nabla f(x_j))^T$$

  with mean $\mathbb{E} \, [\nabla f(x_j) \, (\nabla f(x_j))^T] \;=\; E$

Next: Matrix concentration for $\widehat{E} - E$

Matrix Bernstein Concentration [Minsker 2011, Tropp 2015]

Given

- Independent random symmetric matrices $X_j, \quad 1 \le j \le n$
- Norm: $\max_{1\le j\le n} \|X_j\| \;\le\; \beta$
- Zero mean: $\mathbb{E}\left[X_j\right] = 0, \quad 1 \le j \le n$
- Matrix variance: $\sum_{j=1}^{n} \mathbb{E}\left[X_j^2\right] \;\preceq\; P$ for some P
- Tolerance: $\epsilon \;\ge\; \|P\|^{1/2} + \beta$

Matrix Bernstein Concentration [Minsker 2011, Tropp 2015]

Given

- Independent random symmetric matrices $X_j, \quad 1 \leq j \leq n$
- Norm: $\max_{1 \leq j \leq n} \|X_j\| \leq \beta$
- Zero mean: $\mathbb{E}[X_j] = 0, \quad 1 \leq j \leq n$
- Matrix variance: $\sum_{j=1}^{n} \mathbb{E}[X_j^2] \preceq P$ for some P
- Tolerance: $\epsilon \geq \|P\|^{1/2} + \beta$

Probability that the sum is "large"

$$\mathbb{P}\left[\left\|\sum_{j=1}^{n} X_j\right\| \geq \epsilon\right] \leq 4\,\frac{\text{trace}(P)}{\|P\|}\,\exp\left(\frac{-\epsilon^2/2}{\|P\| + \beta\epsilon/3}\right)$$

Interpretation of Matrix Concentration

$$\mathbb{P}\left[\left\|\sum_{j=1}^{n} X_j\right\| \geq \epsilon\right] \;\leq\; 4\,\frac{\mathrm{trace}\,(P)}{\|P\|}\;\exp\left(\frac{-\epsilon^2/2}{\|P\| + \beta\epsilon/3}\right)$$

Interpretation of Matrix Concentration

$$\mathbb{P}\left[\left\|\sum_{j=1}^{n} X_j\right\| \geq \epsilon\right] \;\leq\; 4\,\frac{\text{trace}\,(P)}{\|P\|}\;\exp\left(\frac{-\epsilon^2/2}{\|P\| + \beta\epsilon/3}\right)$$

- Sum $=$ Deviation from the mean

$$\sum_{j=1}^{n} X_j \;=\; \sum_{j=1}^{n} X_j - \underbrace{\mathbb{E}\left[\sum_{j=1}^{n} X_j\right]}_{=0}$$

- Variance: Numerical rank* of P $=$ Stable rank of $P^{1/2}$

$$\frac{\text{trace}\,(P)}{\|P\|_2} \;=\; \left(\frac{\|P^{1/2}\|_F}{\|P^{1/2}\|_2}\right)^2$$

* Intrinsic dimension, effective rank

Applying the Matrix Concentration

Check the assumptions for $\widehat{E} - E = \sum_{j=1}^{n} X_j$

- Independent random: $X_j \equiv \frac{1}{n}\left(\nabla f(x_j)\,(\nabla f(x_j))^T - E\right)$
- Zero mean: $\mathbb{E}\,[X_j] = 0$
- Bounded norm: $\|X_j\| \leq L^2/n$
- Variance: $\mathbb{E}[X_j^2] = \frac{1}{n^2}\left[\int \left(\nabla f(x)(\nabla f(x))^T\right)^2 \rho(x)dx - E^2\right]$
- Bound for variance: $P = \frac{L^2}{n}E$

$$\int \left(\nabla f(x)(\nabla f(x))^T\right)^2 \rho(x)dx = \underbrace{\|\nabla f(x)\|^2}_{L^2} \underbrace{\int \nabla f(x)(\nabla f(x))^T \rho(x)dx}_{E}$$

Applying the Matrix Concentration

Absolute error:

$$\mathbb{P}\left[\|\widehat{E} - E\| \geq \hat{\epsilon}\right] \;\leq\; 4\,\frac{\operatorname{trace}(E)}{\|E\|}\;\exp\left(-\frac{n}{L^2}\,\frac{\hat{\epsilon}^2/2}{\|E\| + \hat{\epsilon}/3}\right)$$

No explicit dependence on problem dimension m

Error small, if E has low numerical rank

Applying the Matrix Concentration

Absolute error:

$$\mathbb{P}\left[\|\widehat{E} - E\| \geq \hat{\epsilon}\right] \;\leq\; 4\,\frac{\text{trace}\,(E)}{\|E\|}\;\exp\left(-\frac{n}{L^2}\,\frac{\hat{\epsilon}^2/2}{\|E\| + \hat{\epsilon}/3}\right)$$

No explicit dependence on problem dimension m

Error small, if E has low numerical rank

Relative error: Set $\hat{\epsilon} = \|E\|\,\epsilon$

$$\mathbb{P}\left[\frac{\|\widehat{E} - E\|}{\|E\|} \geq \epsilon\right] \;\leq\; 4\,\frac{\text{trace}\,(E)}{\|E\|}\;\exp\left(-n\,\frac{\|E\|}{L^2}\,\frac{\epsilon^2/2}{1 + \epsilon/3}\right)$$

Failure Probability

Given $0 < \epsilon < 1$

$$\mathbb{P}\left[\frac{\|\widehat{E}-E\|}{\|E\|} \geq \epsilon\right] \leq 4 \underbrace{\frac{\text{trace}\,(E)}{\|E\|}}_{\text{numerical rank of } E} \exp\left(-n\,\frac{\|E\|}{L^2}\,\frac{\epsilon^2/2}{1+\epsilon/3}\right)$$

The probability is high that $\widehat{E}$ has relative error ϵ, if

- Function f is smooth: $L^2/\|E\| \approx 1$
- E has low numerical rank: $\text{trace}\,(E)/\|E\| \ll m$

Relative Error for Monte Carlo Approximation

For any $\delta > 0$, with probability at least $1 - \delta$

$$\frac{\|\widehat{E} - E\|}{\|E\|} \leq \gamma + \sqrt{\gamma(\gamma + 6)}$$

where

$$\gamma \equiv \frac{1}{3n} \frac{L^2}{\|E\|} \ln\left(\frac{4}{\delta} \frac{\mathrm{trace}(E)}{\|E\|}\right)$$

With probability $1 - \delta$, approximation $\widehat{E}$ is accurate, if

- Function f is smooth: $L^2/\|E\| \approx 1$
- E has low numerical rank: $\mathrm{trace}(E)/\|E\| \ll m$

Number of Monte Carlo Samples

For any $\delta > 0$, with probability at least $1 - \delta$

$$\frac{\|\widehat{E} - E\|}{\|E\|} \leq \epsilon$$

if number of Monte Carlo samples is

$$n \ \geq \ \frac{3}{\epsilon^2} \, \frac{L^2}{\|E\|} \, \ln\left(\frac{4}{\delta} \, \frac{\mathrm{trace}\,(E)}{\|E\|}\right)$$

With probability $1 - \delta$, only few samples to compute $\widehat{E}$, if

- Function f is smooth: $L^2/\|E\| \approx 1$
- E has low numerical rank: $\mathrm{trace}\,(E)/\|E\| \ \ll \ m$

Final bound: Deterministic + Probabilistic

Assumptions

- Lipschitz constant: $\|\nabla f(x)\| \leq L$
- Eigenvalues of E:

$$\underbrace{\lambda_1 \geq \cdots \geq \lambda_k}_{\text{Active subspace } \mathcal{S}} \quad \underbrace{\gg}_{\text{gap} = \frac{\lambda_k - \lambda_{k+1}}{\lambda_1}} \quad \lambda_{k+1} \geq \cdots \geq \lambda_m \geq 0$$

- Numerical rank: $\text{nr}(E) = (\lambda_1 + \cdots + \lambda_m)/\lambda_1$
- User-specified error tolerance: $0 < \epsilon < \text{gap}/4$
- User-specified failure probability: $0 < \delta < 1$

Number of Monte Carlo Samples for Subspace Approximation

With probability at least $1 - \delta$

$$\sin \angle(\mathcal{S}, \widehat{\mathcal{S}}) \;\leq\; 4\,\epsilon/\text{gap}$$

if number of samples for approximating E is

$$n \;\geq\; \frac{3}{\epsilon^2}\,\frac{L^2}{\|E\|}\,\ln\left(\frac{4}{\delta}\,\text{nr}(E)\right)$$

With high probability, only few samples for accurate subspace $\widehat{\mathcal{S}}$, if

- Function f is smooth: $L^2/\|E\| \approx 1$
- E has low numerical rank: $\text{nr}(E) \;\ll\; m$
- Subspace $\mathcal{S}$ is well-conditioned: $\text{gap} \;\gg\; 1$

Summary

Want: Active subspace $\mathcal{S}$ of function $f : \mathbb{R}^m \rightarrow \mathbb{R}$
Dominant eigenspace of "sensitivity" matrix $E \in \mathbb{R}^{m \times m}$

Compute: Subspace $\widehat{\mathcal{S}}$ from Monte Carlo approximation of E

Contribution: Probabilistic bounds for $\sin \angle(\mathcal{S}, \widehat{\mathcal{S}})$

- No explicit dependence on problem dimension m
- Number of samples to achieve user-specified error at user-specified probability
- Monte Carlo efficient if

  E has low numerical rank
  Subspace $\mathcal{S}$ well-conditioned (large eigenvalue gap)

- Application: Construction of response surfaces

  System of elliptic PDEs, coefficients are log-Gaussian random fields
  Sensitivity matrix E has dimension $m = 3,495$
  Active subspace $\mathcal{S}$ has dimension $k = 10$
  Response surface accurate to 1-2 digits