Active Subspaces

UQ Reading Group

Mario Teixeira Parente

parente@ma.tum.de

TUM

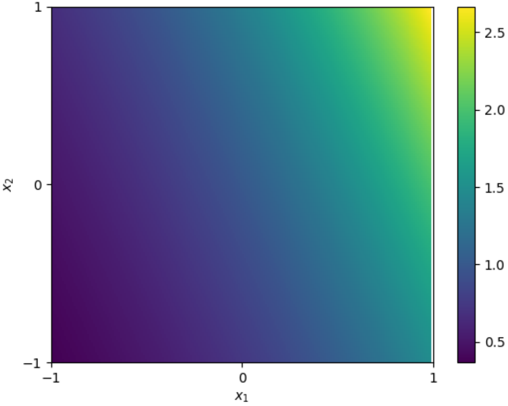
July 11th, 2017

Outline

1. Introductory example
2. Active Subspaces
    - Definition
    - Computation
    - A practical recipe
3. Example: Parametrized PDE

Introductory example

$$f(x) = \exp(0.7x_1 + 0.3x_2)$$



Preparation

1. Normalized inputs $x \in X \subseteq \mathbb{R}^m$
2. Sampling density ρ
3. Largest dimension $k \leq m$
4. Oversampling factor α

Goal

Algorithm

1. Draw $M = \alpha k \log(m)$ independent samples x_j of ρ.
2. Calculate gradient
$$\nabla_x f_j := \nabla_x f(x_j).$$
3. Build
$$\hat{C} = \frac{1}{M} \sum_{j=1}^{M} \nabla_x f_j (\nabla_x f_j)^T = \hat{W} \hat{\Lambda} \hat{W}^T.$$
4. Use *bootstrapping* to assess the estimates' variability.

Outline

1 Introductory example

2 Active Subspaces
- Definition
- Computation
- A practical recipe

3 Example: Parametrized PDE

Simulation map

$$f : X \to \mathbb{R}, \quad X \subseteq \mathbb{R}^m$$

Assumptions:

- Components of x are independent with mean zero and similarly scaled.
- f is differentiable and Lipschitz continuous
  $\implies \|\nabla_x f(x)\| \leq L \;\; \forall x \in X$
- Simulation can produce gradient (through, i.e. adjoint methods)

Definition

$$C = \int_X \nabla_x f (\nabla_x f)^T \rho \, dx$$

Properties:

- $C \in \mathbb{R}^{m \times m}$ is symmetric.
- C is positive semi-definite.

Definition (*cont'd*)

Since C is symmetric, it has a real eigenvalue decomposition

$$C = W \Lambda W^T, \; \Lambda = diag(\lambda_1, \ldots, \lambda_m).$$

Lemma

$$\int_X (\nabla_x f^T w_i)^2 \rho \, dx = \lambda_i$$

Proof.

$\lambda_i = w_i^T C w_i = \int_X w_i^T \nabla_x f (\nabla_x f)^T w_i \rho \, dx = \int_X (\nabla_x f^T w_i)^2 \rho \, dx$ ∎

Definition (*cont'd*)

$$W = \begin{bmatrix} W_1 & W_2 \end{bmatrix}, \ \Lambda = \begin{bmatrix} \Lambda_1 & \\ & \Lambda_2 \end{bmatrix}$$

$$W_1 \in \mathbb{R}^{m \times n}, W_2 \in \mathbb{R}^{m \times (m-n)}$$

$$x = WW^T x = W_1 \underbrace{W_1^T x}_{y} + W_2 \underbrace{W_2^T x}_{z}$$

$$= W_1 y + W_2 z, \quad y \in \mathbb{R}^n, \ z \in \mathbb{R}^{m-n}$$

Definition (*cont'd*)

$$f(x) = f(W_1 y + W_2 z)$$

Lemma

$$\int_X \nabla_y f^T \nabla_y f \rho \, dx = \lambda_1 + \ldots + \lambda_n$$

$$\int_X \nabla_z f^T \nabla_z f \rho \, dx = \lambda_{n+1} + \ldots + \lambda_m$$

$\mathrm{ran}(W_1) \rightarrow$ *active subspace*
$y = W_1^T x \rightarrow$ *active variables*

Do **not** ...

... confuse with the following:

- The active subspace is **not** a subset of X.
- We are **not** restricting our attention to input parameters in the active subspaces.
- We are **not** interested in a low-rank approximation

$$C \approx W_1 \Lambda_1 W_1^T.$$

Instead: Approximate f by a function of fewer than m variables.

Computation

Algorithm

1. Draw M independent samples of ρ.
2. Compute $\nabla_x f_j := \nabla_x f(x_j)$.
3. Build
$$C \approx \hat{C} = \frac{1}{M} \sum_{j=1}^{M} \nabla_x f_j (\nabla_x f_j)^T.$$
4. Compute eigenvalue decomposition
$$\hat{C} = \hat{W} \hat{\Lambda} \hat{W}^T.$$

Computation (*cont'd*)

Theorem

Assume that $\|\nabla_x f(x)\| \leq L \ \forall x \in X$. *Then, for* $\epsilon \in (0, 1]$

$$\mathbb{P}(\hat{\lambda}_k \geq (1+\epsilon)\lambda_k) \leq (m-k+1)\exp\left(-\frac{M\lambda_k\epsilon^2}{4L^2}\right)$$

and

$$\mathbb{P}(\hat{\lambda}_k \leq (1-\epsilon)\lambda_k) \leq k\exp\left(-\frac{M\lambda_k^2\epsilon^2}{4\lambda_1 L^2}\right).$$

Computation (*cont'd*)

Corollary

For $\epsilon \in (0, 1]$,

$$M = \Omega\left(\frac{L^2 \lambda_1}{\lambda_k^2 \epsilon^2} \log(m)\right)$$

implies that $\left|\hat{\lambda}_k - \lambda_k\right| \leq \epsilon \lambda_k$ *with high probability.*

Another result:

Corollary

Let $\epsilon > 0$ *be such that*

$$\epsilon \leq \frac{\lambda_n - \lambda_{n+1}}{5\lambda_1},$$

and choose $M = \Omega\left(\frac{L^2}{\lambda_1 \epsilon^2} \log(m)\right)$*. Then, with high probability*

$$dist\left(ran(W_1), ran(\hat{W}_1)\right) \leq \frac{4\lambda_1 \epsilon}{\lambda_n - \lambda_{n+1}}.$$

Computation (*cont'd*)

Approximate gradients

Assume

$$\|g(x) - \nabla_x f(x)\| \leq \sqrt{m}\gamma_h \quad \forall x \in X$$

with

$$\lim_{h \to 0} \gamma_h = 0.$$

Define

$$G := \int_X gg^T \rho \, dx = U \Theta U^T, \qquad \Theta = \operatorname{diag}(\theta_1, \ldots, \theta_m)$$

$$\hat{G} := \frac{1}{M} \sum_{j=1}^{M} g_j g_j^T = \hat{U} \hat{\Theta} \hat{U}^T, \qquad \hat{\Theta} = \operatorname{diag}(\hat{\theta}_1, \ldots, \hat{\theta}_m).$$

Computation (*cont'd*)

Approximate gradients

Theorem

For $\epsilon \in (0, 1]$, *if* $M = \Omega\left(\frac{L^2\lambda_1}{\lambda_k^2\epsilon^2}\log(m)\right)$, *then*

$$\left|\lambda_k - \hat{\theta}_k\right| \leq \epsilon\lambda_k + \sqrt{m}\gamma_h\left(\sqrt{m}\gamma_h + 2L\right)$$

with high probability.

A practical recipe

Choose M

$$M = \alpha k \log(m)$$

α oversampling factor between 2 and 10
k number of eigenvalues we want to look at
$\log(m)$ from theorems above

A practical recipe (*cont'd*)

Bootstrap

Algorithm

1. Compute gradient examples $\nabla_x f_1, \ldots, \nabla_x f_M$ and estimate $\hat{\Lambda}$ and $\hat{W}$.
2. Choose M_{boot} (number of bootstrap replicates - typically between 100 and 10000).
   For i from 1 to M_{boot}
   a) Draw random integer j_k between 1 and M for $k = 1, \ldots, M$.
   b) Compute
   $$\hat{C}_i^* = \frac{1}{M} \sum_{k=1}^{M} (\nabla_x f_{j_k}) (\nabla_x f_{j_k})^T.$$
   c) Compute eigendecomposition
   $$\hat{C}_i^* = \hat{W}_i^* \hat{\Lambda}_i^* (\hat{W}_i^*)^T.$$
   d) Compute
   $$e_i^* = \text{dist}\left(\text{ran}(\hat{W}_1), \text{ran}(\hat{W}_{i,1}^*)\right).$$

Outline

1 Introductory example

2 Active Subspaces
- Definition
- Computation
- A practical recipe

3 Example: Parametrized PDE

Example: Parametrized PDE

$$\begin{cases} -\nabla \cdot (a\,\nabla_s u) = 1 & \text{in } D := [0,1]^2 \\ u = 0 & \text{on } \Gamma_1 := \Gamma_T \cup \Gamma_L \cup \Gamma_B \\ \frac{\partial u}{\partial n} = 0 & \text{on } \Gamma_2 := D \backslash \Gamma_1 \end{cases}$$

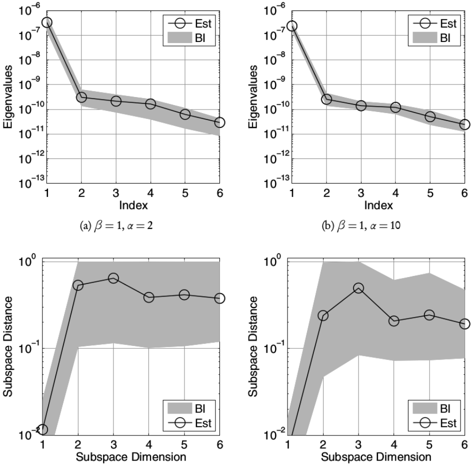
$$\log(a(s,x)) = \sum_{i=1}^{m} x_i \sqrt{\gamma_i} \phi_i(s)$$

(γ_i, ϕ_i) eigenpairs of $\mathcal{C}(s,t) := \exp\left(-\beta^{-1} \|s-t\|_1\right)$

$m = 100$ $(X = \mathbb{R}^{100}), \beta \in \{1, 0.01\}, f(x) := \frac{1}{|\Gamma_2|} \int_{\Gamma_2} u(s,x)\, ds$

Example (*cont'd*)

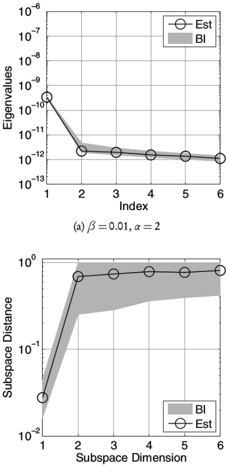
$\beta = 1$



(a) $\beta = 1, \alpha = 2$ (b) $\beta = 1, \alpha = 10$ (c) $\beta = 1, \alpha = 2$ (d) $\beta = 1, \alpha = 10$

Example (*cont'd*)

$\beta = 0.01$



(a) $\beta = 0.01, \alpha = 2$

(b) $\beta = 0.01, \alpha = 10$

(c) $\beta = 0.01, \alpha = 2$

(d) $\beta = 0.01, \alpha = 10$

THE END

Thank you!