SAND2019-4622C

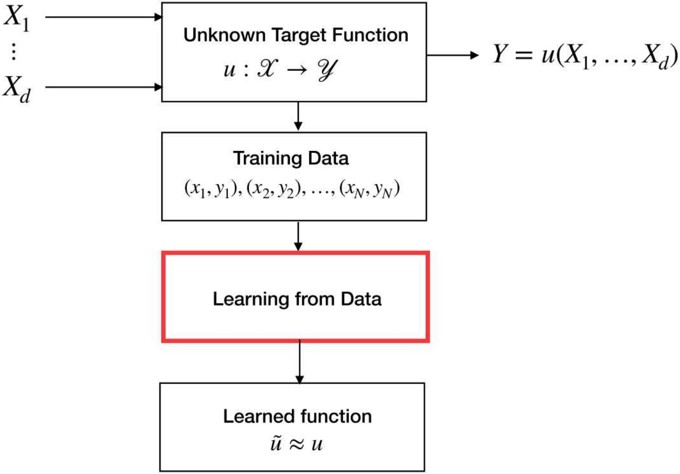
Tensors for Data Science: Concepts, Computing and Applications

P. Rai

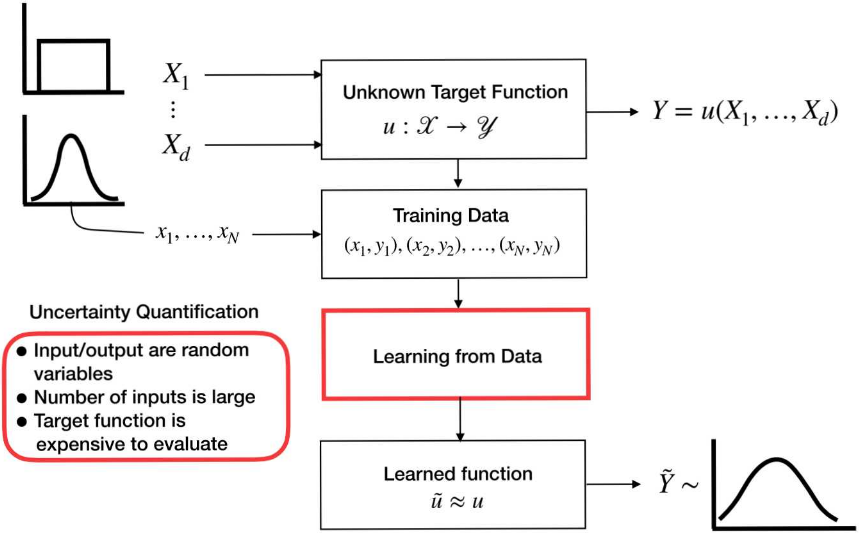
Sandia National Laboratories, California

April 30 2019, IIT Bhilai

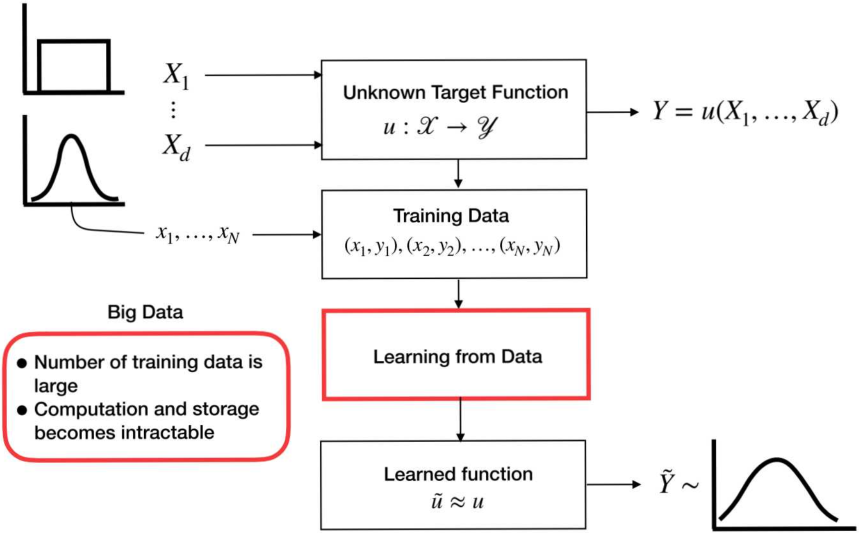
Data Science Framework



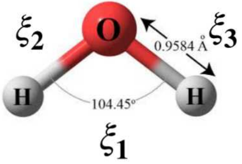
Data Science Framework

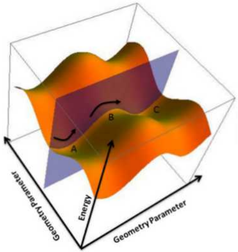


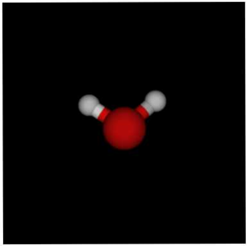
Data Science Framework



Application in Computational Sciences







$$X = \mathbf{A}\xi, \; X_i \sim \mathcal{N}(0,1)$$

Quantity of Interest: $\mathbb{E}(u(X_1, \ldots, X_d))$

- $u(X)$ is anharmonic potential energy surface
- Dimension of $u(X)$: $d = (3 \times \#\text{atoms} - 6)$
  - Water, $d = 3$
  - Benzene, $d = 30$

$\boxed{\text{Approximation of high dimensional functions is difficult}}$

Curse of Dimensionality

Linear approximation

$$u(X) = \sum_{i=1}^{n} u_i \phi_i(X)$$

where $\phi_i(X)$ constitute a suitable basis (ex. polynomial, wavelets etc)

Construction of multidimensional basis

- Choose n basis functions $\{\phi^k_{i_k}\}_{i_k=1}^n$ in dimension $1 \leq k \leq d$

$$u(X_1, \ldots, X_d) = \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} u_{i_1 \ldots i_d} \phi^1_{i_1}(X_1) \ldots \phi^d_{i_d}(X_d), \;\; u_{i_1 \ldots i_d} \in \mathbb{R}$$

- Number of coefficients: $N = n^d$

Can smoothness help?

- If $u \in C^s$, number of coefficients to achieve accuracy ϵ

$$N = O(\epsilon^{-d/s}) \;\; \text{(Curse of dimensionality)}$$

Remedies: Structured approximations

- Low effective dimensionality [1]: $u(X)$ mainly depends on few parameters $X_K,\ K \subset \{1, \ldots, d\}$
- Low-order interactions [2]: Higher order interaction terms have negligible impact
- Sparsity [3]: $u(X)$ is sparse on the basis $\{\phi_i(X), i \in \Lambda\}$

$$u(X) = \sum_{i \in \Lambda_n} u_i \phi_i(X), \Lambda_n \subset \Lambda, \#\Lambda_n = n$$

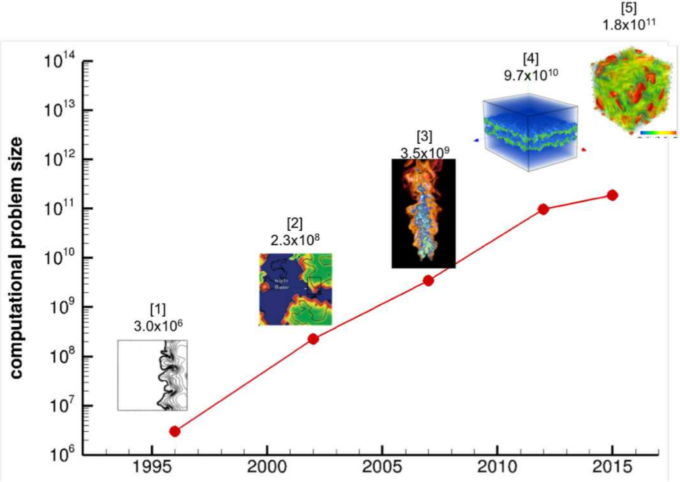
- Low rank [4]: $u(X)$ can be well approximated using separated representation

$$u(X_1, \ldots, X_d) \approx \sum_{i=1}^{r} u_i^1(X_1) \ldots u_i^d(X_d)$$

Key Idea: Combine low rank and sparsity

[1] Constantine '10, Tippireddy'12.
[2] Blatman'09
[3] Doostan'11
[4] Nouy'09

Big (Scientific) Data

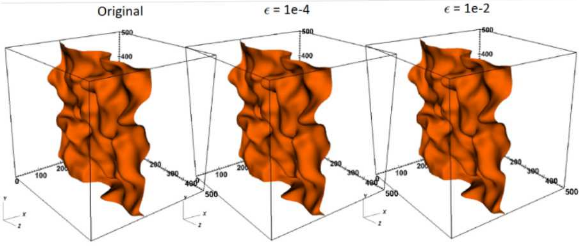


1 T. Echekki, J.H. Chen, Comb. Flame, 1996, vol.106.
2 T. Echekki, J.H, Chen, Proc. Comb. Inst., 2002, vol. 29.
3 R. Sankaran, E.R. Hawkes, J.H. Chen, Proc. Comb. Inst., 2007, vol. 31.
4 E.R. Hawkes, O. Chatakonda, H. Kolla, A.R. Kerstein, J.H. Chen, Comb. Flame, 2012 (online).
5 Gordon Bell submission, 2015

Compression of Big Data

Motivation

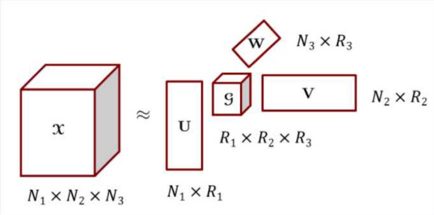
- Data sizes have already outpaced hardware resources
- Compression, even lossy, seems inevitable for
  - Data archival
  - Data analysis
  - Data sharing
- Scientific data usually consist of field data that is smooth and continuous over a domain



400X Compression 20000X Compression

Tensor Based Compression

Compression of Structured Data

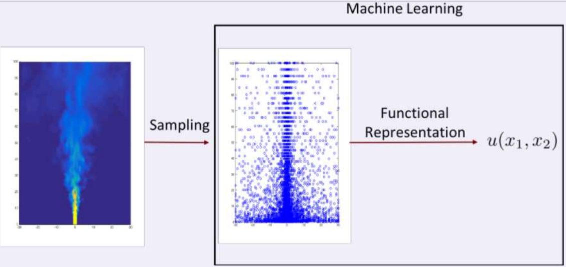


Tucker decomposition [Kolda'17]

$$\mathcal{X} \approx \mathcal{G} \otimes \mathbf{U} \otimes \mathbf{V} \otimes \mathbf{W}$$

$\mathcal{X}$: Structured data tensor
$\mathcal{G}$: Core tensor
$\mathbf{U}, \mathbf{V}, \mathbf{W}$: Factor matrices

Compression of Unstructured Data



Outline

Outline

Canonical rank and tensor format

Canonical rank

- $\mathrm{rank}(u) \leq r : u(X) = \sum_{i=1}^{r} u_i^1(X_1) \dots u_i^d(X_d) = \left(\sum_{i=1}^{r} u_i^1 \otimes \cdots \otimes u_i^d\right)(X)$

Canonical Rank-r

$$\mathcal{R}_r = \left\{ v = \sum_{i=1}^{r} p_i^1 \otimes \ldots \otimes p_i^d \; ; \; p_i^k = \phi(X_k)^T \mathbf{p}_i^k \right\}$$

- Parametrization: $v = F_{\mathcal{R}_r}(\mathbf{p}^1, \ldots, \mathbf{p}^d)$; $\mathbf{p}^k \in \mathbb{R}^{P_k \times r}$

Properties

- $\mathcal{R}_r$ is not closed for $d > 2$ and $r > 1$
- Best approximation problem is not well posed $\longrightarrow$ numerical instabilities

Tree based rank and tensor formats

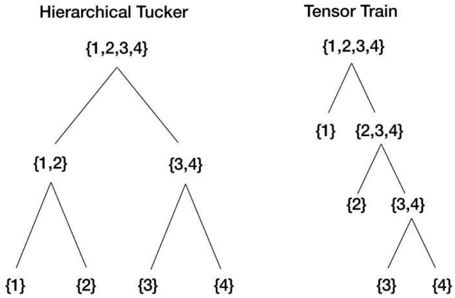
K-rank

- $\mathrm{rank}_K(u) \leq r_K, \; u(X) = \sum_{i=1}^{r_K} u^K(X_K) u^{K^c}(X_{K^c}); \; K \subset \{1, \ldots, d\}$

Tree based rank

- $\mathrm{rank}_K(u) \leq r_K, \; \forall K \in T \subset 2^D$

where T is a tree of dimensions



Tensor Train format

Tensor Train rank

$$\mathcal{TT}_r = \{v : rank_{\{i+1,\dots,d\}}(u) \le r_i, 1 \le i \le d-1\}$$

$$v = \sum_{i_1=1}^{r_1} p^1_{1,i_1} \otimes p^{2,\dots,d}_{i_1}$$

$$v = \sum_{i_1=1}^{r_1} p^1_{1,i_1} \otimes \sum_{i_2=1}^{r_2} p^2_{i_1 i_2} \otimes \dots \otimes \sum_{i_{d-1}=1}^{r_{d-1}} p^{d-1}_{i_{d-2} i_{d-1}} \otimes p^d_{i_{d-1},1}$$

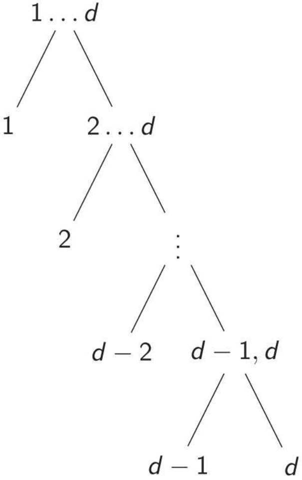
Tensor Train tensor [Oseledets'12]

$$\mathcal{TT}_r = \left\{ v = \sum_{i \in \mathcal{I}} \bigotimes_k p^k_{i_{k-1} i_k};\ p^k_{i_{k-1} i_k} = \phi(X_k)^T \mathbf{p}_{i_{k-1} i_k} \right\}$$

$$\mathcal{I} = \{i = (i_0, i_1, \dots, i_{d-1}, i_d);\ i_k \in \{1,\dots,r_k\}\};\ r_0 = r_d = 1$$

- Parametrization: $v = F_r(\mathbf{p}^1, \dots, \mathbf{p}^d); \mathbf{p}^k = (\mathbb{R}^{P_k})^{r_{k-1} \times r_k}$

Tensor Train format



Tensor Train rank

$$\mathcal{TT}_r = \{v : rank_{\{i+1,\dots,d\}}(u) \le r_i, 1 \le i \le d-1\}$$

$$v = \sum_{i_1=1}^{r_1} p^1_{1,i_1} \otimes p^{2,\dots,d}_{i_1}$$

$$v = \sum_{i_1=1}^{r_1} p^1_{1,i_1} \otimes \sum_{i_2=1}^{r_2} p^2_{i_1 i_2} \otimes \dots \otimes \sum_{i_{d-1}=1}^{r_{d-1}} p^{d-1}_{i_{d-2} i_{d-1}} \otimes p^d_{i_{d-1},1}$$

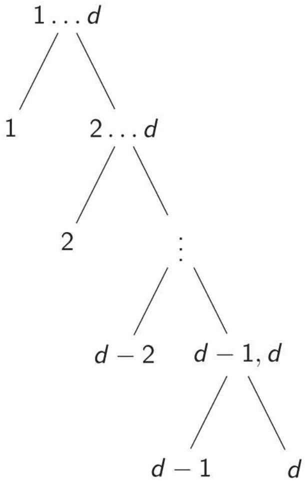
Tensor Train tensor [Oseledets'12]

$$\mathcal{TT}_r = \left\{ v = \sum_{i \in \mathcal{I}} \bigotimes_k p^k_{i_{k-1} i_k};\ p^k_{i_{k-1} i_k} = \phi(X_k)^T \mathbf{p}_{i_{k-1} i_k} \right\}$$

$$\mathcal{I} = \{i = (i_0, i_1, \dots, i_{d-1}, i_d);\ i_k \in \{1, \dots, r_k\}\};\ r_0 = r_d = 1$$

- Parametrization: $v = F_r(\mathbf{p}^1, \dots, \mathbf{p}^d); \mathbf{p}^k = (\mathbb{R}^{P_k})^{r_{k-1} \times r_k}$

Tensor Train format



Tensor Train rank

$$\mathcal{TT}_r = \{v : rank_{\{i+1,\dots,d\}}(u) \leq r_i, 1 \leq i \leq d-1\}$$

$$v = \sum_{i_1=1}^{r_1} p^1_{1,i_1} \otimes p^{2,\dots,d}_{i_1}$$

$$v = \sum_{i_1=1}^{r_1} p^1_{1,i_1} \otimes \sum_{i_2=1}^{r_2} p^2_{i_1 i_2} \otimes \dots \otimes \sum_{i_{d-1}=1}^{r_{d-1}} p^{d-1}_{i_{d-2} i_{d-1}} \otimes p^d_{i_{d-1},1}$$

Tensor Train tensor [Oseledets'12]

$$\mathcal{TT}_r = \left\{ v = \sum_{i \in \mathcal{I}} \bigotimes_k p^k_{i_{k-1} i_k};\ p^k_{i_{k-1} i_k} = \phi(X_k)^T \mathbf{p}_{i_{k-1} i_k} \right\}$$

$\mathcal{I} = \{i = (i_0, i_1, \dots, i_{d-1}, i_d);\ i_k \in \{1, \dots, r_k\}\};\ r_0 = r_d = 1$

- Parametrization: $v = F_r(\mathbf{p}^1, \dots, \mathbf{p}^d); \mathbf{p}^k = (\mathbb{R}^{P_k})^{r_{k-1} \times r_k}$

Low rank formats: Approximation

Minimization problem

$$\min_{\mathbf{p}^1,\dots,\mathbf{p}^d} \|u - F_{\mathcal{M}_r}(\mathbf{p}^1,\dots,\mathbf{p}^d)\|_Q^2 \quad \text{with} \quad \|u-v\|_Q^2 = \sum_{q=1}^{Q} |u(y_q) - v(y_q)|^2$$

Alternating least-squares (ALS)

- For $r \in R$
  - For $1 \leq i \leq d$ and for fixed $\mathbf{p}^j, j \neq i$

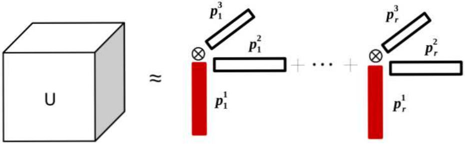
$$\min_{\mathbf{p}^i} \|u - F_{\mathcal{M}_r}(\mathbf{p}^1,\dots,\mathbf{p}^i,\dots,\mathbf{p}^d)\|_Q^2$$

- Select optimal rank in R using cross validation (K-fold)

ALS approximation

Canonical format

- Direct construction: Solve $\min_{v \in \mathcal{R}_r} \|u - v\|_Q^2$
- Greedy construction
  - Set $u_0 = 0$
  - For $r \geq 1$, Solve $\min_{v \in \mathcal{R}_1} \|u - u_{r-1} - v\|_Q^2$



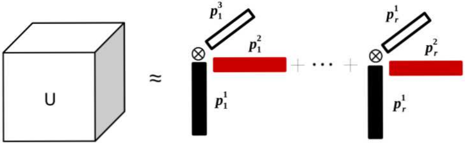
Tensor Train format

- Solve $\min_{v \in \mathcal{TT}_r} \|u - v\|_Q^2$
- Combinatorial problem for the selection of rank

ALS approximation

Canonical format

- Direct construction: Solve $\min_{v \in \mathcal{R}_r} \|u - v\|_Q^2$
- Greedy construction
  - Set $u_0 = 0$
  - For $r \geq 1$, Solve $\min_{v \in \mathcal{R}_1} \|u - u_{r-1} - v\|_Q^2$



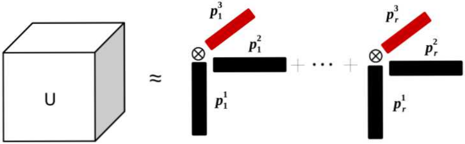
Tensor Train format

- Solve $\min_{v \in \mathcal{TT}_r} \|u - v\|_Q^2$
- Combinatorial problem for the selection of rank

ALS approximation

Canonical format

- Direct construction: Solve $\min_{v \in \mathcal{R}_r} \|u - v\|_Q^2$
- Greedy construction
  - Set $u_0 = 0$
  - For $r \geq 1$, Solve $\min_{v \in \mathcal{R}_1} \|u - u_{r-1} - v\|_Q^2$



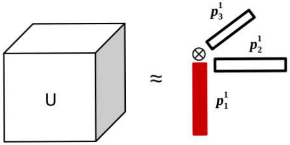
Tensor Train format

- Solve $\min_{v \in \mathcal{TT}_r} \|u - v\|_Q^2$
- Combinatorial problem for the selection of rank

ALS approximation

Canonical format

- Direct construction: Solve $\min_{v \in \mathcal{R}_r} \|u - v\|_Q^2$
- Greedy construction
  - Set $u_0 = 0$
  - For $r \geq 1$, Solve $\min_{v \in \mathcal{R}_{\mathbf{1}}} \|u - u_{r-1} - v\|_Q^2$



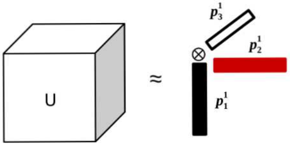
Tensor Train format

- Solve $\min_{v \in \mathcal{TT}_r} \|u - v\|_Q^2$
- Combinatorial problem for the selection of rank

ALS approximation

Canonical format

- Direct construction: Solve $\min_{v \in \mathcal{R}_r} \|u - v\|_Q^2$
- Greedy construction
  - Set $u_0 = 0$
  - For $r \geq 1$, Solve $\min_{v \in \mathcal{R}_\mathbf{1}} \|u - u_{r-1} - v\|_Q^2$



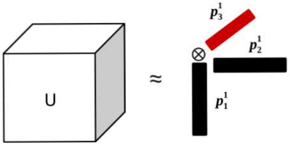
Tensor Train format

- Solve $\min_{v \in \mathcal{TT}_r} \|u - v\|_Q^2$
- Combinatorial problem for the selection of rank

ALS approximation

Canonical format

- Direct construction: Solve $\min_{v \in \mathcal{R}_r} \|u - v\|_Q^2$
- Greedy construction
  - Set $u_0 = 0$
  - For $r \geq 1$, Solve $\min_{v \in \mathcal{R}_{\mathbf{1}}} \|u - u_{r-1} - v\|_Q^2$



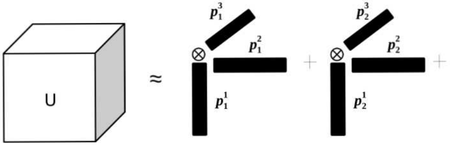
Tensor Train format

- Solve $\min_{v \in \mathcal{TT}_r} \|u - v\|_Q^2$
- Combinatorial problem for the selection of rank

ALS approximation

Canonical format

- Direct construction: Solve $\min_{v \in \mathcal{R}_r} \|u - v\|_Q^2$
- Greedy construction
  - ☐ Set $u_0 = 0$
  - ☐ For $r \geq 1$, Solve $\min_{v \in \mathcal{R}_\mathbf{1}} \|u - u_{r-1} - v\|_Q^2$



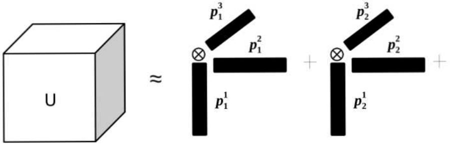
Tensor Train format

- Solve $\min_{v \in \mathcal{TT}_r} \|u - v\|_Q^2$
- Combinatorial problem for the selection of rank

ALS approximation

Canonical format

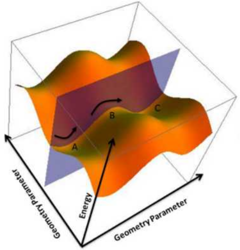
- Direct construction: Solve $\min_{v \in \mathcal{R}_r} \|u - v\|_Q^2$
- Greedy construction
  - Set $u_0 = 0$
  - For $r \geq 1$, Solve $\min_{v \in \mathcal{R}_\mathbf{1}} \|u - u_{r-1} - v\|_Q^2$



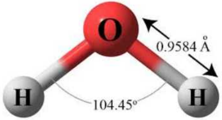
Tensor Train format

- Solve $\min_{v \in \mathcal{TT}_r} \|u - v\|_Q^2$
- Combinatorial problem for the selection of rank

Illustration: Low Rank Approximation of Potential Energy



Potential Energy Surface



Water ($d = 3$)

- Approximation Space: $\otimes_{k=1}^{d} \mathbb{P}_4^k$
- Rank: $1 \leq r \leq 30$
- $\epsilon = \frac{\|u - u_r\|_{Q'}}{\|u\|_{Q'}}$, $Q' = 10000$

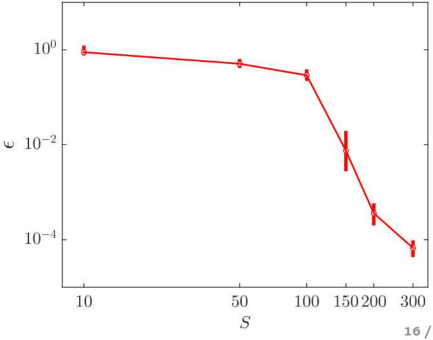
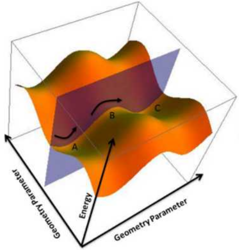
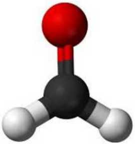


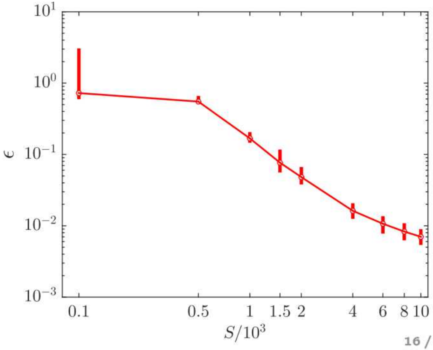
Illustration: Low Rank Approximation of Potential Energy



Potential Energy Surface

Formaldehyde (d = 6)





- Approximation Space: $\bigotimes_{k=1}^{d} \mathbb{P}_4^k$
- Rank: $1 \leq r \leq 30$
- $\epsilon = \frac{\|u - u_r\|_{Q'}}{\|u\|_{Q'}}, \; Q' = 10000$

Sampling and Computational Complexity

Sampling Complexity

- Approximation of multivariate functions in polynomial spaces [Chkifa'13]

$$Q \sim (\#\Lambda)^2; \quad \Lambda : \text{Basis set}$$

- Question: Number of samples w.r.t number of parameters in low rank format?

$$Q \sim f(d, r, n)$$

- Conjecture: For canonical rank-r format in polynomial spaces [Rai '15]
  - Greedy: $Q \sim dr(n)^2$
  - Direct: $Q \sim d(nr)^2$

Computation Complexity

- Complexity of (Direct) ALS: $\mathcal{O}(r^2 dn^2 Q) \to \mathcal{O}(r^3 d^2 n^4)$
- Complexity of (Greedy) ALS: $\mathcal{O}(dn^2 Q) \to \mathcal{O}(rd^2 n^4)$

Can we further reduce the number of samples?

Outline

Sparsity in linear approximation

Least-squares with sparse regularization

$$u(X) = \sum_{i=1}^{P} v_i \phi_i(X) = \phi(X)^T \mathbf{v}$$

- If $u(X)$ is sparse on the basis set $\{\phi_i\}_{i=1}^P$

$$\min_{\mathbf{v}\in\mathbb{R}^P} \|\mathbf{u} - \mathbf{\Phi v}\|_2^2 \; s.t. \; \|\mathbf{v}\|_0 \leq m \quad \text{or} \quad \min_{\mathbf{v}\in\mathbb{R}^P} \|\mathbf{u} - \mathbf{\Phi v}\|_2^2 + \lambda\|\mathbf{v}\|_0$$

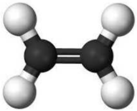
- Convexification

$$\min_{\mathbf{v}\in\mathbb{R}^P} \|\mathbf{u} - \mathbf{\Phi v}\|_2^2 + \lambda\|\mathbf{v}\|_1$$

Selection of λ

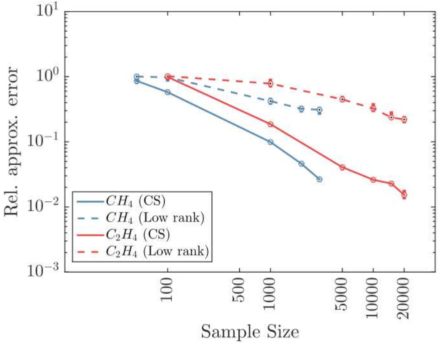
- Least Angle Regression for solving sparse least-squares [Efron'04]
- Fast leave one out cross validation for selection of optimal sparse vector

Illustration: Sparse approximation of Potential Energy



Approximation Space: $\otimes_{k=1}^{9}\mathbb{P}_4^k$
Number of basis = 1953125

Approximation Space: $\otimes_{k=1}^{12}\mathbb{P}_4^k$
Number of basis = 244140625



Approximation in sparse tensor formats

Sparse low rank tensor formats $\mathcal{M}^{\text{m-sparse}}$

$$\mathcal{M}^{\text{m-sparse}} = \{v = F_{\mathcal{M}}(\mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^d);\ \|\mathbf{p}^k\|_0 \leq m_k;\ 1 \leq k \leq d\}$$

Alternating least-squares with sparse regularization

- ALS with sparse regularization

$$\min_{\mathbf{p}^1, \dots, \mathbf{p}^d} \|u - F_{\mathcal{M}}(\mathbf{p}^1, \dots, \mathbf{p}^d)\|_Q^2 \ \ s.t.\ \ \|\mathbf{p}^k\|_0 \leq m_k$$

Approximation in sparse tensor formats

Sparse low rank tensor formats $\mathcal{M}^{\text{m-sparse}}$

$$\mathcal{M}^{\text{m-sparse}} = \{v = F_{\mathcal{M}}(\mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^d);\ \|\mathbf{p}^k\|_0 \leq m_k;\ 1 \leq k \leq d\}$$

Alternating least-squares with sparse regularization

- ALS with sparse regularization

$$\min_{\mathbf{p}^1,\dots,\mathbf{p}^d} \|u - F_{\mathcal{M}}(\mathbf{p}^1, \dots, \mathbf{p}^d)\|_Q^2 + \sum_{k=1}^{d} \lambda_k \|\mathbf{p}^k\|_1$$

Approximation in sparse tensor formats

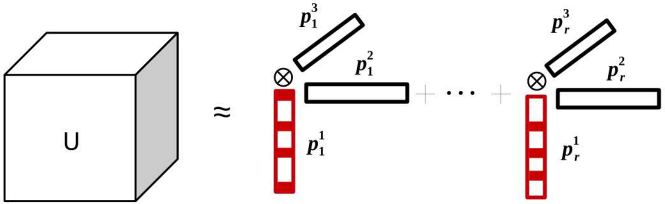
Sparse low rank tensor formats $\mathcal{M}^{\text{m-sparse}}$

$$\mathcal{M}^{\text{m-sparse}} = \{v = F_{\mathcal{M}}(\mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^d);\ \|\mathbf{p}^k\|_0 \leq m_k;\ 1 \leq k \leq d\}$$

Alternating least-squares with sparse regularization

- ALS with sparse regularization

$$\min_{\mathbf{p}^1,\dots,\mathbf{p}^d} \|u - F_{\mathcal{M}}(\mathbf{p}^1, \dots, \mathbf{p}^d)\|_Q^2 + \sum_{k=1}^{d} \lambda_k \|\mathbf{p}^k\|_1$$



Approximation in sparse tensor formats

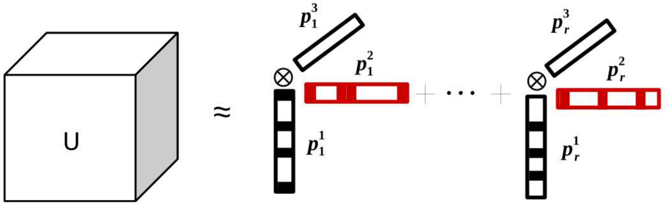
Sparse low rank tensor formats $\mathcal{M}^{\text{m-sparse}}$

$$\mathcal{M}^{\text{m-sparse}} = \{v = F_{\mathcal{M}}(\mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^d);\ \|\mathbf{p}^k\|_0 \le m_k;\ 1 \le k \le d\}$$

Alternating least-squares with sparse regularization

- ALS with sparse regularization

$$\min_{\mathbf{p}^1,\dots,\mathbf{p}^d} \|u - F_{\mathcal{M}}(\mathbf{p}^1, \dots, \mathbf{p}^d)\|_Q^2 + \sum_{k=1}^{d} \lambda_k \|\mathbf{p}^k\|_1$$



Approximation in sparse tensor formats

Sparse low rank tensor formats $\mathcal{M}^{\text{m-sparse}}$

$$\mathcal{M}^{\text{m-sparse}} = \{v = F_{\mathcal{M}}(\mathbf{p}^1, \mathbf{p}^2, \dots, \mathbf{p}^d);\ \|\mathbf{p}^k\|_0 \le m_k;\ 1 \le k \le d\}$$

Alternating least-squares with sparse regularization

- ALS with sparse regularization

$$\min_{\mathbf{p}^1,\dots,\mathbf{p}^d} \|u - F_{\mathcal{M}}(\mathbf{p}^1, \dots, \mathbf{p}^d)\|_Q^2 + \sum_{k=1}^{d} \lambda_k \|\mathbf{p}^k\|_1$$

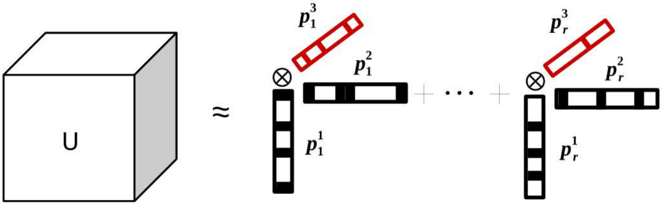
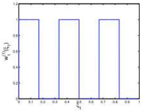


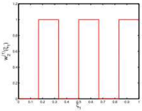
Illustration:Checkerboard function

Rank-2 function:

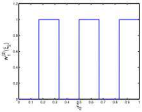
$$u(X_1, X_2) = \sum_{i=1}^{2} w_i^{(1)}(X_1) w_i^{(2)}(X_2)$$



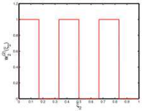
(a) $w_1^{(1)}(X_1)$



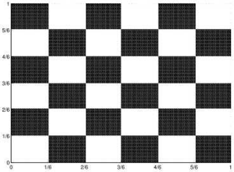
(b) $w_2^{(1)}(X_1)$



(c) $w_1^{(2)}(X_2)$



(d) $w_2^{(2)}(X_2)$



$X \backsim \mathcal{U}(0,1)$

Approximation of u in $\mathcal{S}^1_{P_1} \otimes \mathcal{S}^2_{P_2}$

Piecewise polynomials of degree p defined on a uniform partition of s intervals:

$$\boxed{\mathcal{S}^k_{P_k} = \mathbb{P}_{p,s}}$$

Illustration: Checkerboard function

Checkerboard function ($Q = 200$)

	Sparse	Low rank		Sparse low rank	
Approximation space	Error	Error	Rank	Error	Rank
$\mathbb{P}_{2,6} \otimes \mathbb{P}_{2,6}$, $P = 18^2$	0.5591	0.664	2	$2.41\,10^{-13}$	2
$\mathbb{P}_{2,12} \otimes \mathbb{P}_{2,12}$, $P = 36^2$	0.9867	-	-	$1.50\,10^{-12}$	2
$\mathbb{P}_{10,6} \otimes \mathbb{P}_{10,6}$, $P = 66^2$	1.6575	-	-	$7.88\,10^{-13}$	2

With few samples:

- Sparse low rank gives quasi-exact recovery
- Rank 2 is retrieved

ALS in Tensor Train Format

For $1 \leq k \leq d-1$

- Compute sparse $w^{k,k+1}$ by solving

$$\min_{\mathbf{w}^{k,k+1}} \|\mathrm{u} - \mathrm{F}_r^k(\mathbf{p}_1, \ldots, \mathbf{w}^{k,k+1}, \ldots, \mathbf{p}^d)\|_Q^2 + \lambda_k \|vec(\mathbf{w}^{k,k+1})\|_1$$

- Compute truncated low-rank approximation of $w^{k,k+1} \approx \sum_{i_k=1}^{r_k^*} p_{i_{k-1} i_k}^k \otimes p_{i_k i_{k+1}}^{k+1}$ using SVD $\rightarrow$ adaptive rank r_k^* selected using cross validation

$$v = \sum_{i_1=1}^{r_1} \ldots \sum_{i_k=1}^{r_k^*} \ldots \sum_{i_{d-1}=1}^{r_{d-1}} p_{1 i_1}^1 \otimes \ldots \otimes p_{i_{k-1} i_k}^k \otimes p_{i_k i_{k+1}}^{k+1} \otimes \ldots \otimes p_{i_{d-1} 1}^d$$

Illustration for TT approximation: Borehole Function

Borehole function models water flow through a borehole:

$$u(X) = \frac{2\pi X_3(X_4 - X_6)}{\log(\frac{r(X_2)}{x_1})(1 + \frac{2X_7X_3}{\log(\frac{r(X_2)}{X_1})X_1^2X_8} + \frac{X_3}{X_5})}$$

$$X = (r_w, r, T_u, H_u, T_l, H_l, L, K_w)$$

r_w	$\mathcal{N}(0.1, 0.0161812)$	Radius of borehole
r	$\mathcal{N}(0, 1)$	Radius of influence
T_u	$\mathcal{U}(63070, 115600)$	Transmissivity of upper aquifer
H_u	$\mathcal{U}(990, 1110)$	Potentiometric head of upper aquifer
T_l	$\mathcal{U}(63.1, 116)$	Transmissivity of lower aquifer
H_l	$\mathcal{U}(700, 820)$	Potentiometric head of lower aquifer
L	$\mathcal{U}(1120, 1680)$	Length of boreholee
K_w	$\mathcal{U}(9855, 12045)$	Hydraulic conductivity of borehole

Illustration for TT approximation: Borehole Function

Table: Borehole function. Relative test error of canonical representation in $\mathcal{R}_r$ and TT representation in $\mathcal{TT}_r$ with isotropic rank and rank adaptation with training sample sets of sizes $Q = 50, 10^2, 10^3, 10^4$.

	format	rank	test error
Q=50	Linear		$8.6\,10^{-2}$
	Canonical	3	$2.6\,10^{-3}$
	TT / isototropic rank	(2 2 2 2 2 2 2)	$3.0\,10^{-3}$
	TT / adapted rank	(2 4 4 2 2 1 1)	$5.9\,10^{-4}$
Q=100	Linear		$1.4\,10^{-2}$
	Canonical	7	$5.4\,10^{-4}$
	TT / isototropic rank	(2 2 2 2 2 2 2)	$5.1\,10^{-4}$
	TT / adapted rank	(2 3 3 2 3 1 1)	$4.9\,10^{-4}$
Q=1000	Linear		$2.2\,10^{-4}$
	Canonical	12	$5.1\,10^{-5}$
	TT / isototropic rank	(3 3 3 3 3 3 3)	$6.8\,10^{-6}$
	TT / adapted rank	(4 3 3 4 4 2 2)	$5.5\,10^{-7}$
Q=10000	Canonical	7	$3.6\,10^{-5}$
	TT / isototropic rank	(5 5 5 5 5 5 5)	$7.5\,10^{-7}$
	TT / adapted rank	(2 2 4 4 4 3 3)	$2.9\,10^{-7}$

Illustration: High Dimensional Integration

Quantity of Interest: $\mathbb{E}(u(X_1, \ldots, X_d))$

Sparse low rank approximation: $u(X_1, \ldots, X_d) \approx \sum_{i=1}^{r} \prod_{k=1}^{d} v_i^k(X_k)$

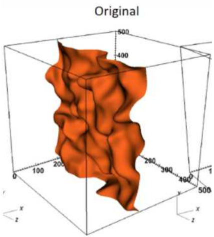
$$\mathbb{E}(u) \approx \sum_{i=1}^{r} \prod_{k=1}^{d} \mathbb{E}(v_i^k(X_k))$$

Molecule	Sparsity+Low Rank	Low Rank	Monte Carlo
H_2O $(d = 3)$	100	150	7×10^5
H_2CO $(d = 6)$	450	4000	-
CH_4 $(d = 9)$	3000	-	-
C_2H_4 $(d = 12)$	12000	-	-

Samples for $< 1\mathrm{cm}^{-1}$ error on energy and frequency corrections

Outline

Tensors for Big Data Compression



Grid Size:
$500 \times 500 \times 500 \times 400$
(400GB)

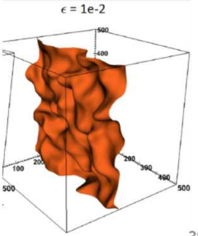
Sampling based surrogate construction $\longrightarrow$

$$u(X) \approx \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \alpha_{i_1 \ldots i_d} \prod_{k=1}^{d} \psi_{i_k}^{k}(X_k)$$

$\#$ Parameters $= 494800$ (3.95 MB)

$$u(X) \approx \sum_{i_1=1}^{r_1} \cdots \sum_{i_d=1}^{r_d} \alpha_{i_1 \ldots i_d} \prod_{k=1}^{d} \psi_{i_k}^{k}(X_k)$$

Surrogate evaluation $\longrightarrow$



Randomized Least Squares

Minimization Problem

$$\min_{v\in\mathbb{R}^n}\|\mathbf{\Phi v}-\mathbf{u}\|_2, \mathbf{\Phi}\in\mathbb{R}^{Q\times n}, Q\gg n$$

$$\downarrow$$

$$\min_{v\in\mathbb{R}^n}\|\mathbf{M\Phi v}-\mathbf{Mu}\|_2, \mathbf{M}\in\mathbb{R}^{S\times n}, S\ll Q$$

Leverage Scores

$$\mathbf{\Phi}=\sum_{j=1}^{Q}\sigma_i\mathbf{u}_i\otimes\mathbf{v}_i$$

Leverage score of $\mathbf{\Phi}(i,:)=\|\mathbf{u}_i\|_2^2$

Fast Johnson Lindenstrauss Transformation

$$\min_{v\in\mathbb{R}^n}\|\mathbf{\Phi}^\star\mathbf{v}-\mathbf{u}^\star\|_2$$

$$\mathbf{\Phi}^\star=\mathbf{S}\mathcal{F}\mathbf{D\Phi}$$

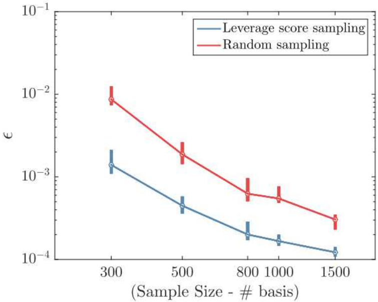
- Diagonal Flip: $\mathbf{D}, \mathbf{D}\in\mathbb{R}^{n\times n}, diag(\mathbf{D})=\pm1$
- Fast Mixing: $\mathcal{F}$
- Uniform row sampling: $\mathbf{S}$

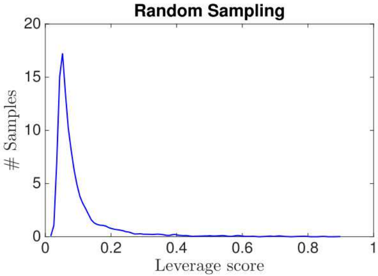
Illustration: Randomized Least Squares

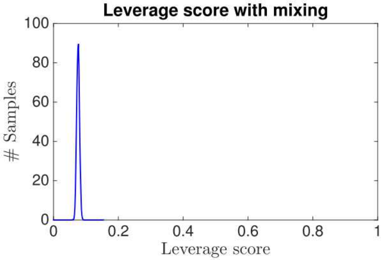
Ishigami Function

$$u(X) = \sin(X_1) + 7\sin^2(X_2) + 0.1X_3^4\sin(X_1)$$
$$X_i \sim \mathcal{U}[-\pi, \pi]$$

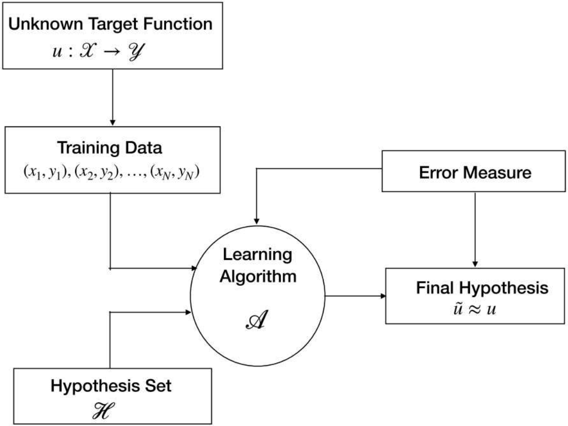
Approximation space: $\otimes_{k=1}^{3}\mathbb{P}_1^{(k)}2$



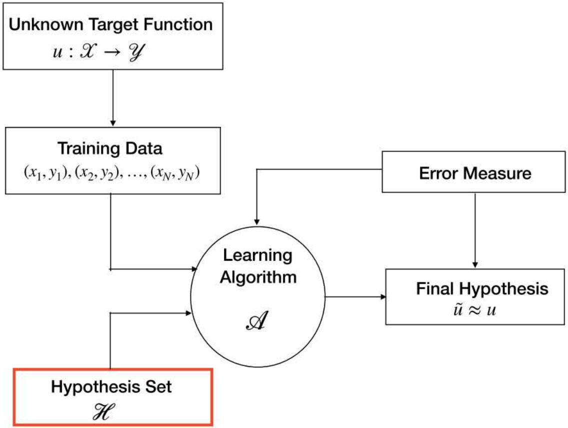




Data Science Framework

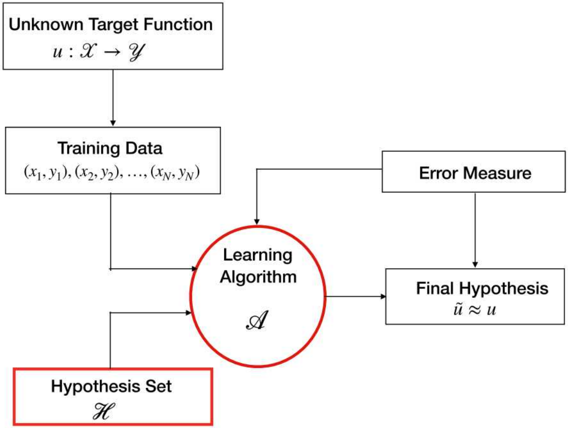


Data Science Framework



- **Sparse Canonical Tensor**
- **Sparse Tensor Train**

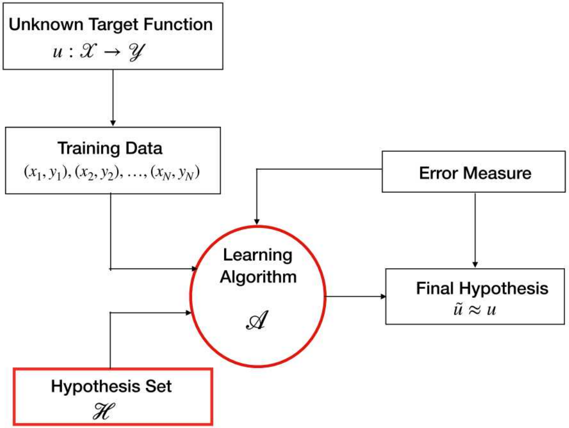
Data Science Framework



- Sparse Canonical Tensor
- Sparse Tensor Train

- Sparse Alternating Least Squares
- Randomized Least Square

Data Science Framework



- Sparse Canonical Tensor
- Sparse Tensor Train

- Sparse Alternating Least Squares
- Randomized Least Square

- Uncertainty Quantification
- Data Compression
- Computational Sciences

Outline

Clustering and Classification

- Objective: Piecewise low rank approximation using domain decomposition for irregular functions
- Joint work with O. Le Maitre (CNRS) and O. Knio (Duke)
- Consider a function that has 'domain wise' low rank approximation

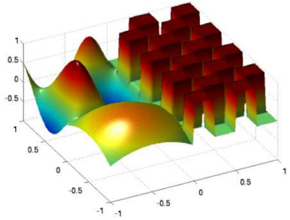
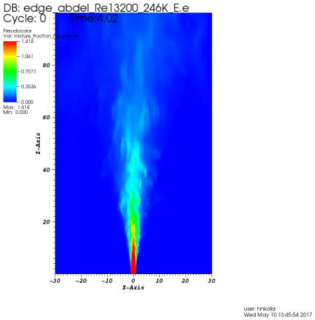


Figure: Manhattan function



Geometric Clustering

Clustering algorithm

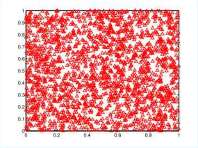
- Sample the initial sample set $\mathcal{K}$
- Repartition the samples in K clusters $\mathcal{K}_k$, $1 \leq k \leq K$ using $K-$means clustering based on minimizing overall geometric intracluster distance

$$\min \sum_{k=1}^{K} G^2(\mathcal{K}_k)$$

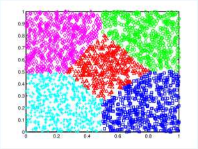
$$G^2(\mathcal{K}_k) = \sum_{q \in \mathcal{K}_k} \|y_q - \bar{y}(\mathcal{K}_k)\|^2$$

$$\bar{y}(\mathcal{K}_k) = \frac{1}{\#\mathcal{K}_k} \sum_{q \in \mathcal{K}_k} y_q$$

Initial Samples



After Clustering



Clustering Scheme

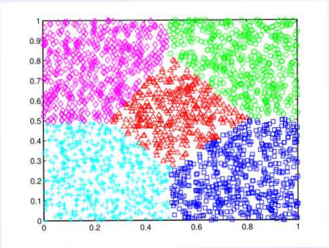
- Fit a low rank model $v_{\mathcal{K}_k}$ in each cluster
- Solve

$$\min \sum R(\mathcal{K}_k) = \sum_{q \in \cup_{k=1}^{K} \mathcal{K}_k} |u(y_q) - v_{\mathcal{K}_k}(y_q)|^2 + \gamma \|y_q - \bar{y}(\mathcal{K}_k)\|^2$$

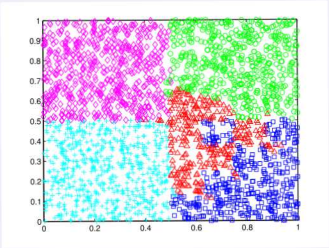
over the set of clusters

- $|u(y_q) - v_{\mathcal{K}_k}(y_q)|$ is estimated using cross validation if $q \in \mathcal{K}_k$

Before minimization



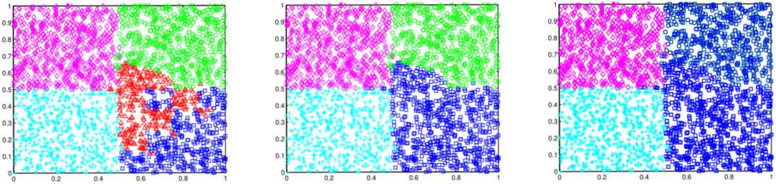
After minimization



Merging Scheme

For adjacent clusters k, k'

- Compute $\Delta(k, k') = R(\mathcal{K}_k \cup \mathcal{K}_{k'}) - R(\mathcal{K}_k) - R(\mathcal{K}_{k'})$; $k, k' \in \mathcal{K}$
- Merge if $\Delta(k, k') < 0$

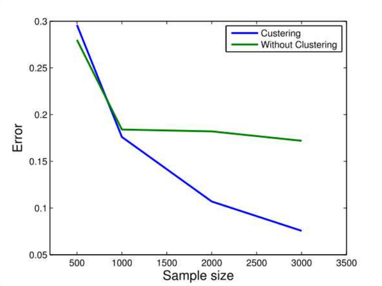


Classification

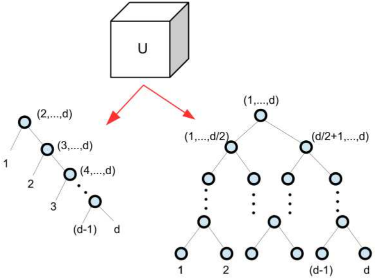
Global approximation using nearest neighbour search

- How to decide the cluster of a new point y?
- Determine the cluster membership of n nearest neighbours of y
- Set the cluster membership of y equal to maximum cluster membership amongst n neighbours

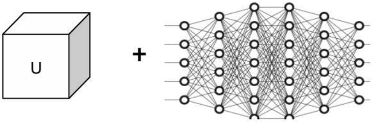
Error v/s Sample size (using Canonical low rank model)



Research Directions



- Choice of new hypothesis sets (Hierarchical Tucker)
- New optimization algorithms (e.g. Stochastic Gradient Descent)



- Tensor based compression of hyper parameters in deep networks
- Connections between mathematical structures of tensors and Neural Networks

Teaching

Contribution to Existing Courses

- BTech Courses
  - Linear Algebra I and II (IC104 and IC152)
  - Probability and Statistics (IC105)
- MTech Courses
  - Linear Algebra (MA501)
  - Numerical Techniques (MA507)
  - Numerical Methods and Computing (MA601)

Introduction of New Courses

- BTech Courses
  - Introduction to Data Science
  - Algorithms in Machine Learning
- MTech Courses
  - Uncertainty Quantification in Predictive Computing
  - Multilinear Algebra for Data Science

Acknowledgement

- Support for this work was provided through the Scientific Discovery through Advanced Computing (SciDAC) program funded by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research.
- Computing resources for this work was provided by DoE's NERSC supercomputing facility.
- Sandia National Laboratories is a multimission laboratory operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration. Sandia has major research and development responsibilities in nuclear deterrence, global security, defense, energy technologies and economic competitiveness, with main facilities in Albuquerque, New Mexico, and Livermore, California.

Papers

[1] Rai P., Sargsyan K., Najm H., Hirata S., *Sparse low rank approximation of potential energy surfaces with applications in estimation of anharmonic zero point energies and frequencies", arXiv:1808.02922 (Submitted to J. Comp. Chem.)*

[2] Rai P., Sargsyan K., Najm H., Hirata S., *Compressed sparse tensor based quadrature for vibrational quantum mechanics integrals", Computer Methods in Applied Mechanics and Engineering, 2018: 336, 471-484*

[3] Rai P., Sargsyan K., Najm H., Hermes M., Hirata S.,*Low-rank canonical-tensor decomposition of potential energy surfaces: Application to grid-based diagrammatic vibrational Green's function theory", Molecular Physics, 2017: (1-15)*

[4] Chevreuil M., Lebrun R., Nouy A., Rai P.,*A least-squares method for sparse low rank approximation of multivariate functions", SIAM/ASA Journal on Uncertainty Quantification, 3.1 (2015): 897-921 (Author list in alphabetical order)*

[5] Rai P.,*Randomized canonical polyadic tensor approximation for quantum mechanics integrals" (Preprint available)*

[6] Rai P., Giraldi L., Nouy A., Chevreuil M.,*Learning algorithms for low rank approximation of multivariate functions in tensor train format" (Preprint available)*

.