Computationally-Efficient Approximations to Arbitrary Linear Dimensionality Reduction Operators

Jarvis Haupt

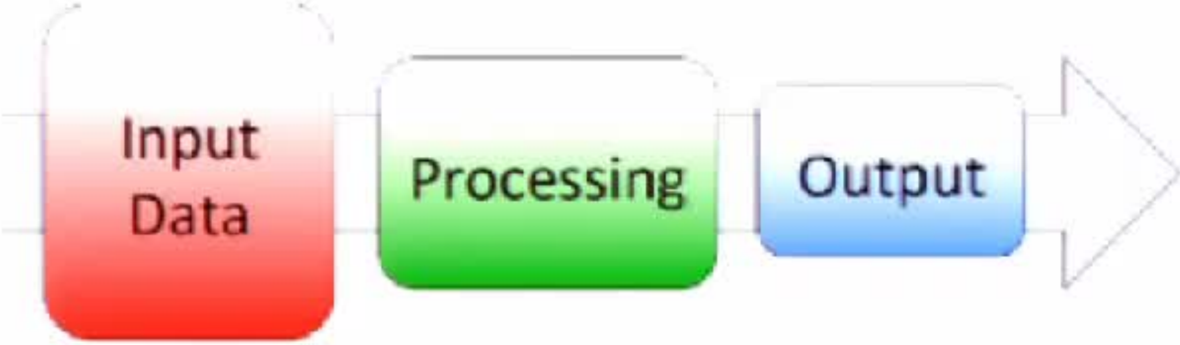
Department of Electrical and Computer Engineering
University of Minnesota

SIAM Conference on Applied Linear Algebra
October 26, 2015

Motivation

large-scale data processing

The (big) data processing pipeline, simplified:



Dilemma: There is an ongoing, inherent *stress* between processing operations that we *desire* to perform, and those we can *tractably* implement!

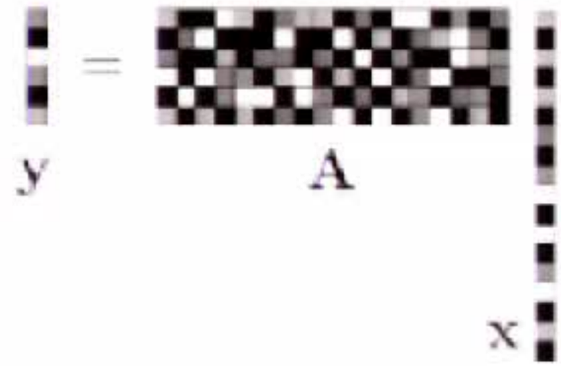
This talk examines (in part) fundamental relationships between these "classes" of operations...

Problem Statement

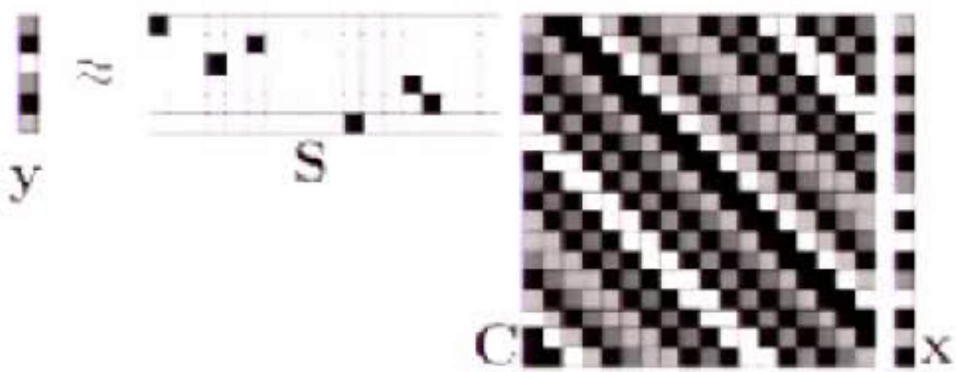
"computationally-efficient" approximations

We consider a common computational "primitive" arising in data processing:

Matrix-vector multiplication with (dimensionality reducing) operator



Our (Initial) Aim:
Approximate original operator using a "partial circulant" operator:



some preliminaries

Let $\mathcal{C}_n$ denote the set of all (real) $n \times n$ circulant matrices of the form

$$\mathbf{C} = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ c_n & c_1 & \cdots & c_{n-1} \\ & & \ddots & \\ c_2 & c_3 & \cdots & c_1 \end{bmatrix},$$

where $\mathbf{c} = [c_1 \ \cdots \ c_n]^T \in \mathbb{R}^n$.

Let $\mathcal{S}_m$ be the set of $m \times n$ ($m < n$) row sampling matrices whose rows comprise m different canonical basis vectors of $\mathbb{R}^n$, with permutations.

The set of all $m \times n$ real *partial circulant* matrices is

$$\mathcal{PC}_{m,n} = \left\{ \mathbf{SC} \in \mathbb{R}^{m \times n} : \mathbf{S} \in \mathcal{S}_m, \mathbf{C} \in \mathcal{C}_n \right\}.$$

A Fundamental "Partial Circulant" Matrix Approximation Result

partial circulant approximation - further insights

Consider matrices **A** w/ row spaces distributed uniformly at random on $\mathrm{Gr}(m, n)$ (Grassmannian manifold of m-dimensional linear subspaces of $\mathbb{R}^n$).

$\Rightarrow$ Quantify *proportion* of matrices w/ accurate partial circulant approximations...

Swayambhoo Jain
(UMN PhD Student)

Thm - Approximable Proportion: **(Swayambhoo Jain & JH 2015)**

For $2 \leq m \leq n$, let $\mathbf{A} \in \mathbb{R}^{m \times n}$ have iid $\mathcal{N}(0, 1)$ entries. For $\delta \in [0, 1/8)$, and n is sufficiently large, there exists a positive constant $c(\delta)$ such that

$$\Pr(\mathcal{E}_{\mathcal{PC}_{m,n}}(\mathbf{A}) < \delta \|\mathbf{A}\|_F^2) = \mathcal{O}(e^{-c(\delta) \cdot mn}).$$

(Submitted; preprint at arxiv:1502.07317)

Take-away: Most (fat) matrices cannot be approximated to high accuracy (in Frob. norm) by partial circulant matrices – the fraction admitting "good" approximations is *exponentially small* in the product of the matrix dimensions.

two extensions

We consider two extensions of the *partial circulant* framework...

1) Post-processing: Rather than approximate $\mathbf{A} \approx \mathbf{SC}$, consider

$$\mathbf{A} \approx \mathbf{PSC},$$

where $\mathbf{S} \in \mathbb{R}^{m' \times n}$ $(m' \geq m)$, $\mathbf{P} \in \mathbb{R}^{m \times m'}$ is *arbitrary* "post-processing" matrix.

two extensions

We consider two extensions of the *partial circulant* framework...

1) Post-processing: Rather than approximate $\mathbf{A} \approx \mathbf{SC}$, consider

$$\mathbf{A} \approx \mathbf{PSC},$$

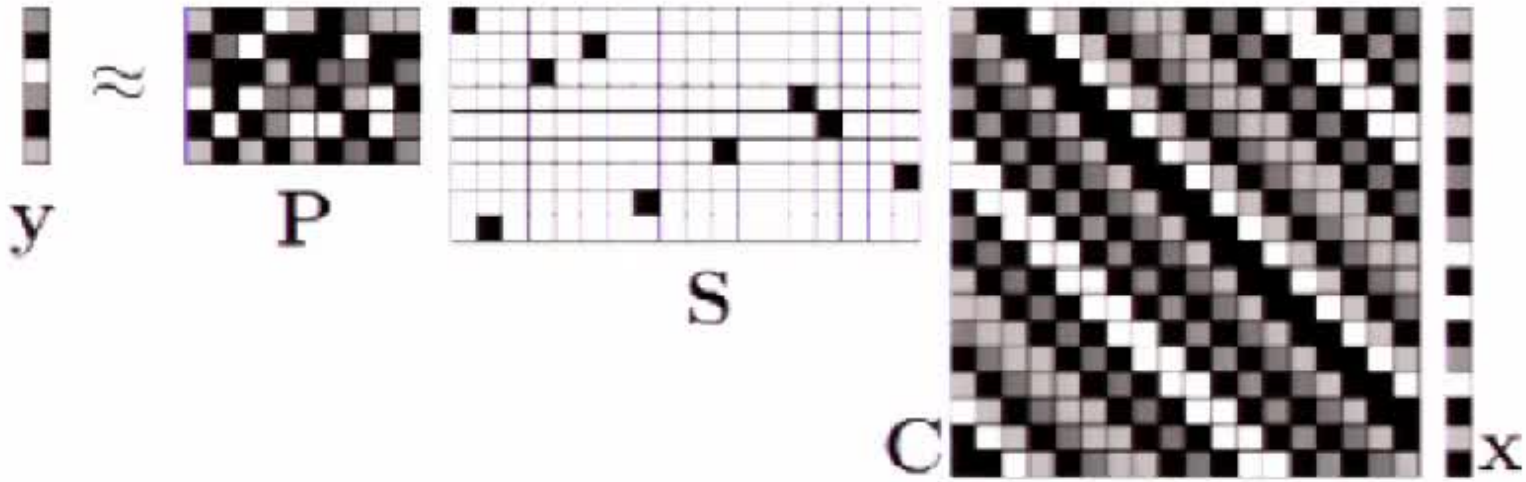
where $\mathbf{S} \in \mathbb{R}^{m' \times n}$ $(m' \geq m)$, $\mathbf{P} \in \mathbb{R}^{m \times m'}$ is *arbitrary* "post-processing" matrix.

Implications:

- Enables accurate approximations when row space of **A** (not **A** itself) well approximated by vectors related by circular shifts
- Slightly higher computational complexity $\to \mathcal{O}(mm' + n \log n)$
  (can still be $\mathcal{O}(n \log n)$ when m' is small wrt n, e.g., when $m' = o(n^{1/2})$...)

graphically...

Generalized approach → approximations of the form



two extensions

2) Restricted Input Domains: Approximate "action" of **A** for vectors **x** belonging to "restricted" set $\mathcal{X}$ of inputs (e.g., a subspace, union of subspaces, manifold...)

New $\mathcal{X}$-dependent approximation metrics, e.g., for any loss/distortion function $\ell : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{R}^+$, can consider the "worst-case" distortion

$$\sup_{\mathbf{x} \in \mathcal{X}} \ell(\mathbf{A}\mathbf{x}, \mathbf{P}\mathbf{S}\mathbf{C}\mathbf{x}),$$

or "average case" distortion

$$\mathbb{E}_{\mathbf{x} \sim p_{\mathcal{X}}} \left[\ell(\mathbf{A}\mathbf{x}, \mathbf{P}\mathbf{S}\mathbf{C}\mathbf{x}) \right],$$

for a specified distribution $p_{\mathcal{X}}$ defined on $x \in \mathcal{X}$.

Implications:

- leverages/utilizes application domain knowledge
- doesn't require approximating **A** outside of "interesting" inputs $\in \mathcal{X}$

experimental investigation

Consider 1440 vectorized, resized (to 45×45) images from *COIL-20* image database (http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php)

Let rows of **A** (to be approximated) be the top 30 principal component vectors

Here: Minimize an (empirical) average Frobenius approximation error...let **X** be a matrix whose columns are the vectorized images described above. Then, seek to minimize

$$\|\mathbf{AX} - \mathbf{PSCX}\|_F$$

by choice of matrices **P**, **S**, and **C**.

algorithmic approach

Algorithm 1 "Data-Driven" Partial Circulant Approximation

Inputs: LDR matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, parameters $\lambda, \mu, \epsilon > 0$,
Matrix of "representative" data $\mathbf{X} \in \mathbb{R}^{n \times p}$.
Initialize: $\mathbf{M}^{(0)} = \mathbf{U}\Sigma$ (from the SVD $\mathbf{AX} = \mathbf{U}\Sigma\mathbf{V}^T$), $\text{obj}^{(0)} = \|\mathbf{AX}\|_F^2$
repeat
$\mathbf{C}^{(t)} = \arg\min_{\mathbf{C} \in \mathcal{C}_n} \|\mathbf{AX} - \mathbf{M}^{(t-1)}\mathbf{CX}\|_F^2 + \mu\|\mathbf{C}\|_F^2$
$\mathbf{M}^{(t)} = \arg\min_{\mathbf{M} \in \mathbb{R}^{m \times n}} \|\mathbf{AX} - \mathbf{MC}^{(t)}\mathbf{X}\|_F^2 + \lambda\|\mathbf{M}\|_{2,1}$
$\text{obj}^{(t)} = \|\mathbf{AX} - \mathbf{M}^{(t)}\mathbf{C}^{(t)}\mathbf{X}\|_F^2 + \mu\|\mathbf{C}^{(t)}\|_F^2 + \lambda\|\mathbf{M}^{(t)}\|_{2,1}$
until $\text{obj}^{(t)} - \text{obj}^{(t-1)} \leq \epsilon \cdot \text{obj}^{(t-1)}$
Output: $\mathbf{M}^* = \mathbf{M}^{(t)}, \mathbf{C}^* = \mathbf{C}^{(t)}$

Here, $\|\mathbf{M}\|_{1,2} = \sum_{j=1}^{n} \|\mathbf{M}_{:,j}\|_2$, where $\|\mathbf{M}_{:,j}\|_2$ is Euclidean norm of column $\mathbf{M}_{:,j}$.

Insight: Combine actions of **P** and **S** into **M**; enforce **M** to be *column-sparse*!

analysis

We have established the following preliminary result for this general framework:

Theorem: **(Swayambhoo Jain & JH 2015)**

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ *be any fixed matrix, and* $\ell : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ *any loss that is L-Lipschitz continuous (in Frobenius norm). Let* $\mathcal{X}$ *be any countable set of n-dimensional unit-norm vectors.*

For any $\epsilon \in (0, 1/2)$, *there exists a post-processing* $\mathbf{P} \in \mathbb{C}^{m \times m'}$, *sampling matrix* $\mathbf{S} \in \mathbb{R}^{m' \times n}$ *comprised of rows of identity, and circulant* $\mathbf{C} \in \mathbb{C}^{n \times n}$ *for which*

$$\sup_{\mathbf{x} \in \mathcal{X}} \ell(\mathbf{Ax}, \mathbf{PSCx}) \leq L\epsilon \|\mathbf{A}\|_F,$$

provided that

$$m' > c_1 \epsilon^{-2} \log(c_2 m |\mathcal{X}|) \log^4(n).$$

Here, c_1 *and* c_2 *are universal positive constants.*

Utilizes intermediate results of Lap, Wakin, & Rozell 2013 and stable embedding results from Davenport et al. 2010

Extensions to uncountable sets $\mathcal{X}$ can be derived using *covering* arguments...

Summary & Acknowledgments