SAND2014-15533PE

Illustration by Chris Brigman

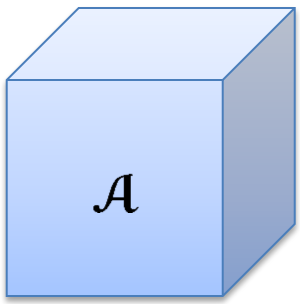
Real-valued Symmetric Tensor Decompositions

Tamara G. Kolda
Sandia National Labs
Livermore, CA

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Symmetric Tensors

- $\mathbb{S}^{[m,n]}$ = set of m-way, n-dimensional real-valued symmetric tensors
- m = number of modes or ways
- n = size of each mode
- symmetric = entries invariant to permutation of indices



3-way tensor

Symmetry for 3-way tensor (m = 3)

$$a_{ijk} = a_{ikj} = a_{jik} = a_{kij} = a_{jki} = a_{kji}$$
$$\text{for all } i, j, k \in \{1, 2, \ldots, n\}$$

- Applications of symmetric tensors: diffusion tensor imaging (DTI/HARDI), higher-order statistics, higher-order derivatives, relativity, signal processing, etc.

Tensor Norm

Given an m-way tensor of dimension n

We use boldface $\mathbf{i}$ to denote a multi-index

$$\mathbf{i} = (i_1, i_2, \ldots, i_m)$$

The norm is the square root of the sum of squares of the elements

$$\|\mathcal{A}\|^2 = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_m=1}^{n} a_{\mathbf{i}}^2$$

$$\|\mathcal{A} - \mathcal{B}\|^2 = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_m=1}^{n} (a_{\mathbf{i}} - b_{\mathbf{i}})^2$$

Rank-1 Symmetric Tensors

Consider a real-valued n-vector

$$\mathbf{x} \in \mathbb{R}^n$$

The m-way outer product is denoted by

$$\mathbf{x}^m \equiv \mathbf{x} \circ \mathbf{x} \circ \cdots \circ \mathbf{x}$$

We use boldface **i** to denote a multi-index

$$\mathbf{i} = (i_1, i_2, \ldots, i_m)$$

Entry **i** of the outer product is defined by

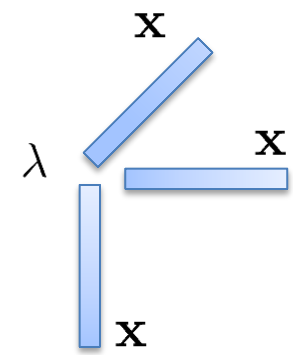
$$(\mathbf{x}^m)_{\mathbf{i}} = x_{i_1} x_{i_2} \cdots x_{i_m}$$

A rank-one tensor has the form

$$\lambda \mathbf{x}^m$$

3-way, rank-1 tensor

$$\lambda \mathbf{x}^m \equiv \lambda \, \mathbf{x} \circ \mathbf{x} \circ \mathbf{x}$$

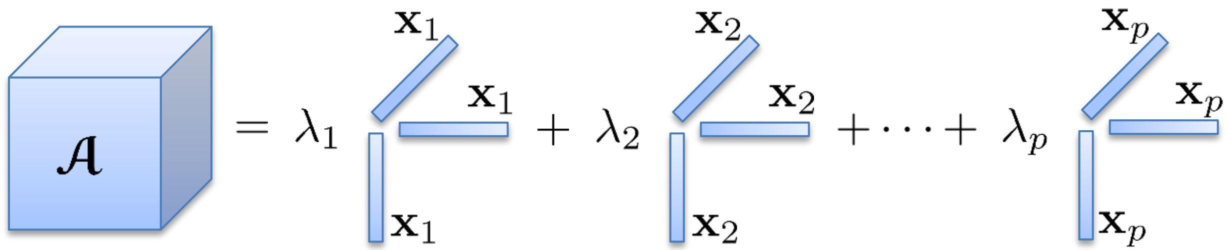


Note: λ only has to be separate when m is even.

Symmetric Tensor Decomposition

From Comon et al. (2008), any real-valued symmetric tensor can be decomposed as the sum of symmetric rank-1 outer products:

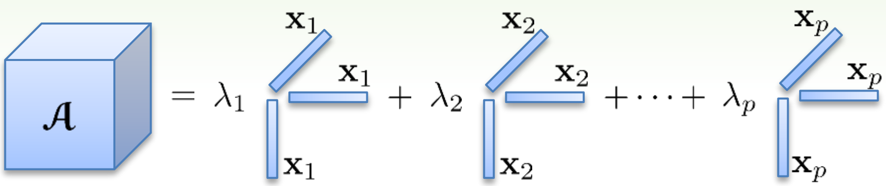
$$\mathcal{A} = \sum_{k=1}^{p} \lambda_k \, \mathbf{x}_k^m$$

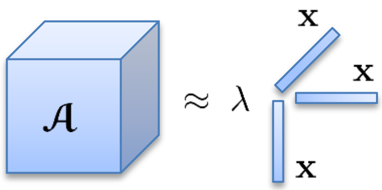


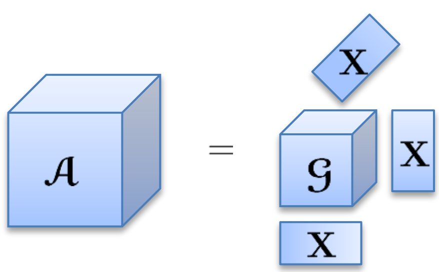
Goal: Find $\boldsymbol{\lambda} \in \mathbb{R}^n$ and $\mathbf{X} \in \mathbb{R}^{n \times p}$ (assuming p given) where
λ_k denotes the kth entry of the vector $\boldsymbol{\lambda}$ and
$\mathbf{x}_k$ denotes the kth column of the matrix $\mathbf{X}$

Unknown p is a topic of future work.

Related Work

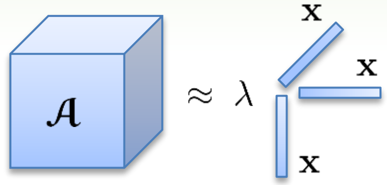






- **Numerical approaches**
  - Shashua, Zass, Hazan (2006): N-way nonnegative clustering using sampling and only indices such that $i_1 < i_2 < \cdots < i_m$
  - Cichocki at al. (2007): Nonnegative factorization of sparse symmetric tensors based on KL divergence
  - Wang and Qi (2007): Successive deflation
- **Algebraic geometry approaches**
  - Based on Waring decomposition, but numerical viability is uncertain
  - Brachat et al. (2010)
  - Oeding & Ottaviani (2011)
- **Best symmetric rank-1 approximation**
  - Equivalent to finding the largest magnitude eigenvalue (see next slide)
- **Symmetric tucker approximation**
  - Cambre, De Lathauwer, De Moor (1991)
  - Regalia (2013)

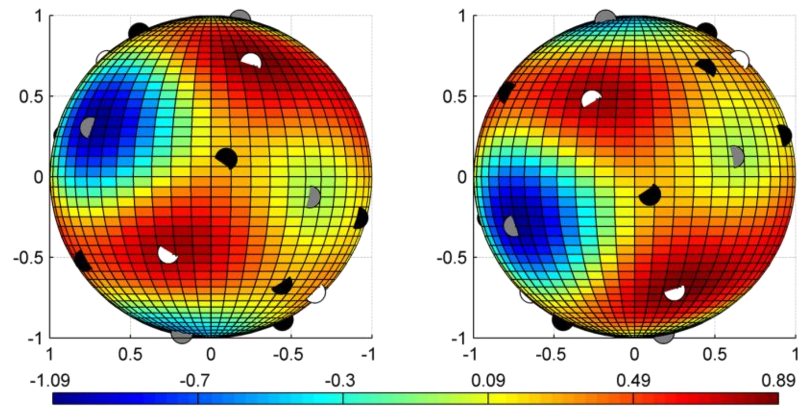
Special case: Symmetric Rank-1 Approximation



$$\min \|\mathcal{A} - \lambda \mathbf{x}^m\|^2$$

- Analogous to the problem of computing the largest Z-eigenpair
- Related Work
  - De Lathauwer et al. (1999): power method
  - Kofidis & Regalia (2002), Regalia & Kofidis (2003): power method analyzed and refined
  - Qi (2005) / Lim (2005): Z-eigenpair defintion
  - Kolda & Mayo (2011), Kolda & Mayo (2014): shifted power method with adaptive shift
  - Han (2012): Optimization formulation
  - Cui, Dai, Nie (2014): SDP formulation

Optimization landscape for particular problem with m=4, n=3



Many extreme points!

White = Local Max

Grey = Local Min

Black = Saddle

Optimization Formulations

Formulation 1: All entries least squares difference

$$f_1(\boldsymbol{\lambda}, \mathbf{X}) = \left\| \mathcal{A} - \sum_{k=1}^{p} \lambda_k \, \mathbf{x}_k^m \right\|^2 = \sum_{i_1, i_2, \cdots, i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \, (\mathbf{x}_k^m)_{\mathbf{i}} \right)^2$$

Problem: Repeated entries are counted multiple times in the summation, giving up to $m!$ more weight to some differences than others.

For m=3 and n=2, element (1,1,1) is only counted once whereas element (1,1,2) = (1,2,1) = (2,1,1) is counted three times!

Formulation 2: Distinct entries least squares difference

$$f_2(\boldsymbol{\lambda}, \mathbf{X}) = \sum_{i_1 \leq i_2 \leq \cdots \leq i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \, (\mathbf{x}_k^m)_{\mathbf{i}} \right)^2$$

Refining the Formulation

$$f_1(\boldsymbol{\lambda}, \mathbf{X}) = \sum_{i_1, i_2, \cdots, i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \left(\mathbf{x}_k^m \right)_{\mathbf{i}} \right)^2 \qquad f_2(\boldsymbol{\lambda}, \mathbf{X}) = \sum_{i_1 \le i_2 \le \cdots \le i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \left(\mathbf{x}_k^m \right)_{\mathbf{i}} \right)^2$$

Scaling Ambiguity: $\mathbf{x}_k \leftarrow \beta \mathbf{x}_k \quad \text{and} \quad \lambda_k \leftarrow \lambda_k / \beta^m$

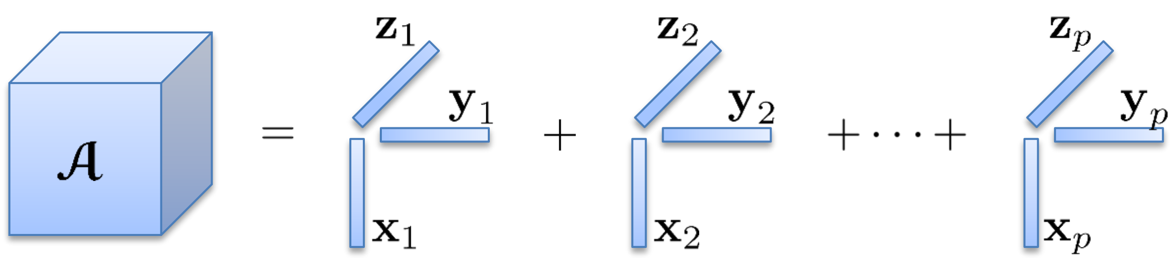
Constraint: $\|\mathbf{x}_k\| = 1 \quad \text{for} \quad k = 1, \ldots, p$

Penalty: $p(\mathbf{X}; \gamma) = \gamma \sum_{k=1}^{p} (\mathbf{x}_k^{\mathsf{T}} \mathbf{x}_k - 1)^2$

Ignoring Symmetry

- What if we just ignore symmetry and use a standard method for tensor factorization?

$$\mathcal{A} = \sum_{k=1}^{p} \mathbf{x}_k \circ \mathbf{y}_k \circ \mathbf{z}_k$$



- Sidiropoulos and Bro (2000): Solution is *essentially* unique if (specialized to symmetric case)

$$2p + (m-1) \leq m \cdot \text{k-rank}(\mathbf{X})$$

- $\Rightarrow$ Ignoring symmetry is okay if there is a unique solution

Experimental Setup

- Five Methods
  - Poblano: $f_1(\boldsymbol{\lambda}, \mathbf{X}) + p(\mathbf{X};\lambda)$
  - SNOPT: $f_1(\boldsymbol{\lambda}, \mathbf{X}) + p(\mathbf{X};\lambda)$
  - Poblano: $f_2(\boldsymbol{\lambda}, \mathbf{X}) + p(\mathbf{X};\lambda)$
  - SNOPT: $f_2(\boldsymbol{\lambda}, \mathbf{X}) + p(\mathbf{X};\lambda)$
  - ALS
- Four Cases
  - Case 1: m = 4, n = 2, p = 2
  - Case 2: m = 4, n = 4, p = 5
  - Case 3: m = 3, n = 10, p = 2
  - Case 4: m = 4, n = 5, p = 3

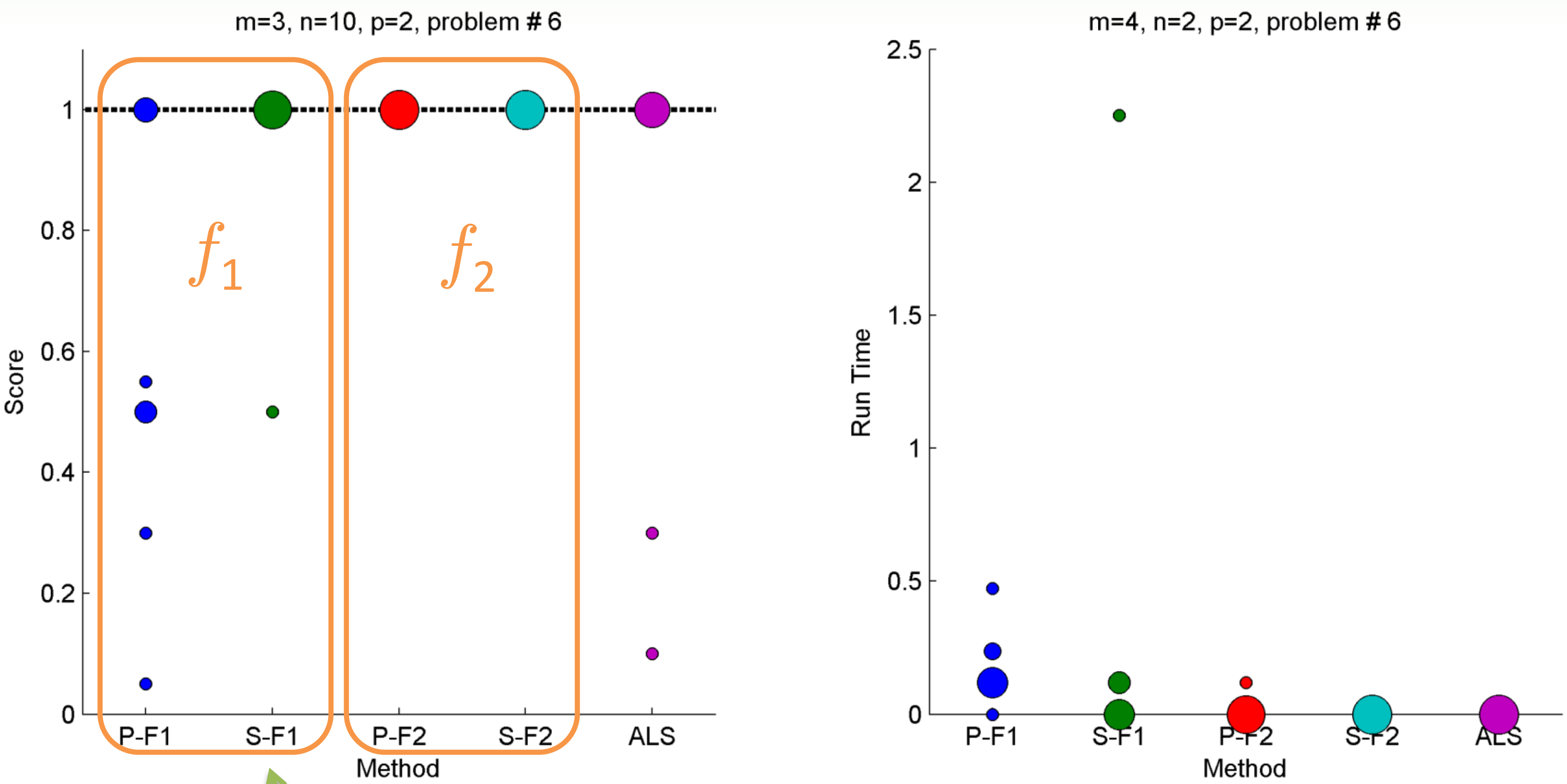
$$f_1(\boldsymbol{\lambda}, \mathbf{X}) = \sum_{i_1,i_2,\cdots,i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \left(\mathbf{x}_k^m\right)_{\mathbf{i}} \right)^2$$

$$f_2(\boldsymbol{\lambda}, \mathbf{X}) = \sum_{i_1 \leq i_2 \leq \cdots \leq i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \left(\mathbf{x}_k^m\right)_{\mathbf{i}} \right)^2$$

$$p(\mathbf{X};\gamma) = \gamma \sum_{k=1}^{p} (\mathbf{x}_k^{\mathsf{T}} \mathbf{x}_k - 1)^2$$

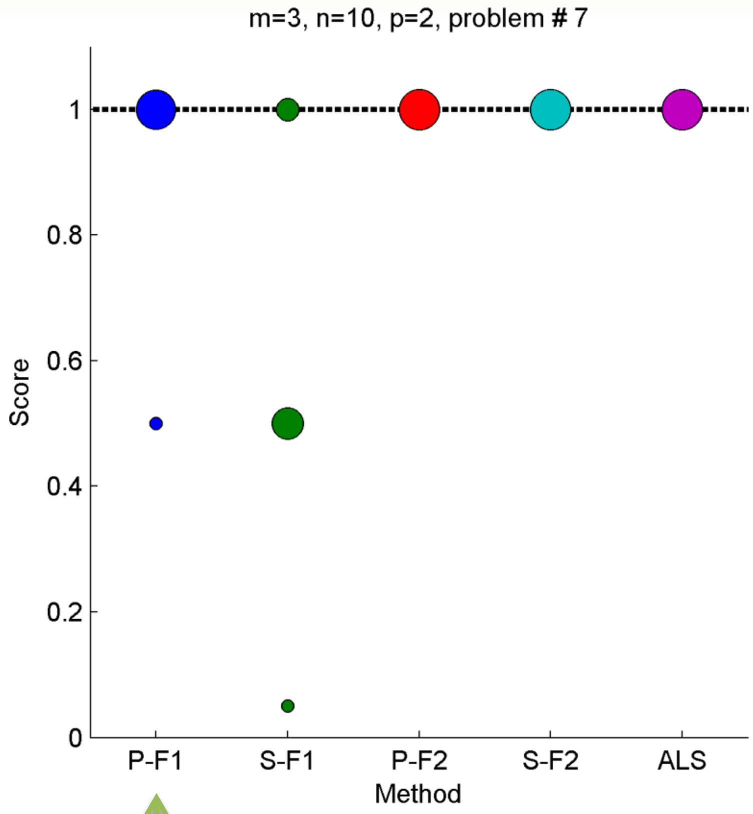
- Optimization Methods
  - Poblano Toolbox for MATLAB
    - L-BFGS
    - Dunlavy, Kolda, Acar (2010)
  - SNOPT Ver. 7 (via MATLAB interface)
    - Sequential quadratic programming
    - Gill, Murray, Saunders (2008)
  - Alternating least squares
    - Solves unsymmetric problem
    - Carroll & Chang (1970), Harshman (1970)
    - Implementation from Tensor Toolbox for MATLAB, Version 2.5
- Creating test problems
  - Choose entries of n x p matrix $\mathbf{X}$ from N(0,1) via `randn` and normalize each column so that $\|\mathbf{x}_k\|$ = 1 for k =1,…,p
  - Choose λ_k = +1 or -1 for k =1,…,p (randomly)

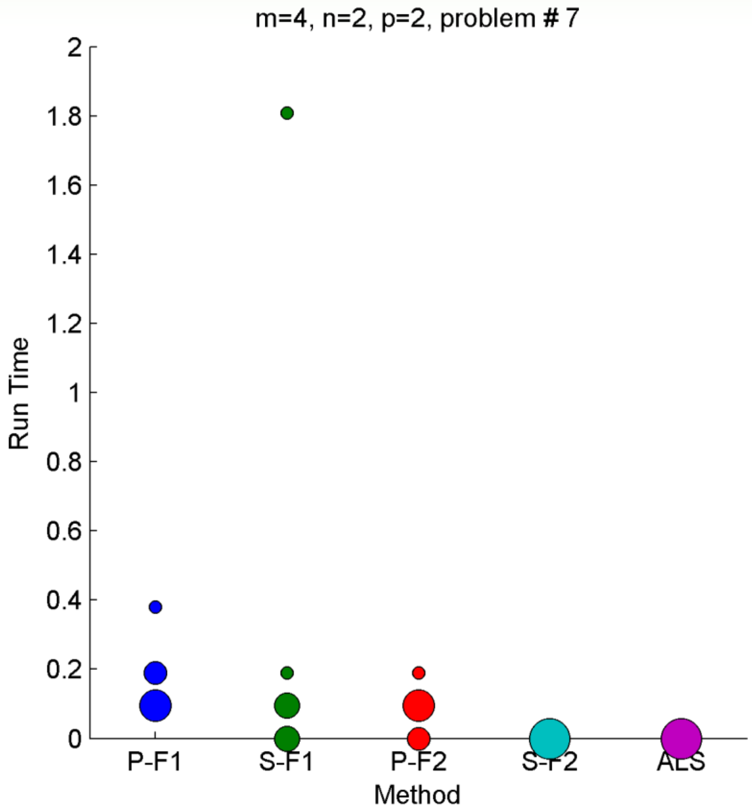
Sample Results: f_2 better than f_1



Aside: SNOPT solves more problems for f_1

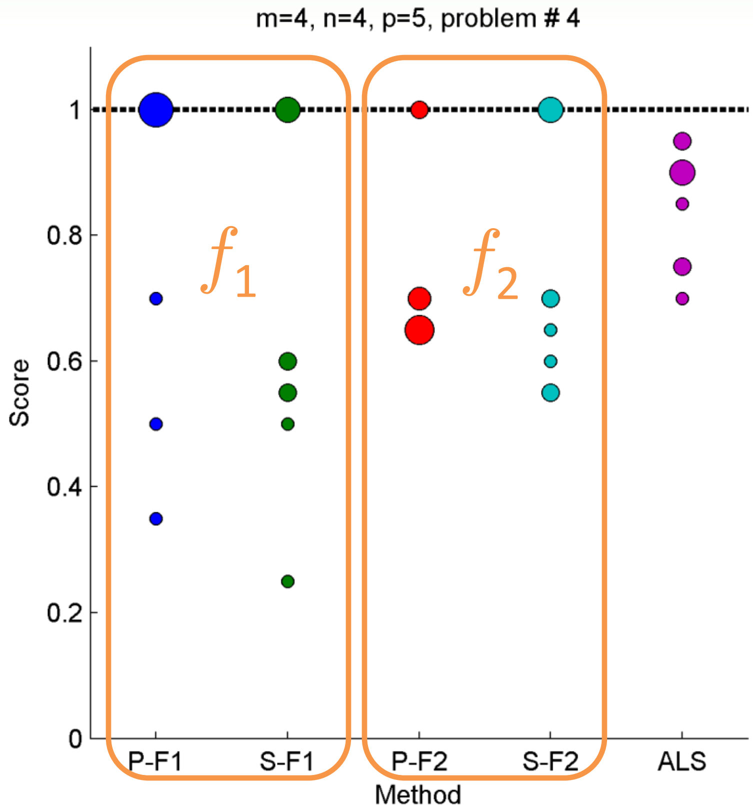
Sample Results: ALS may do well

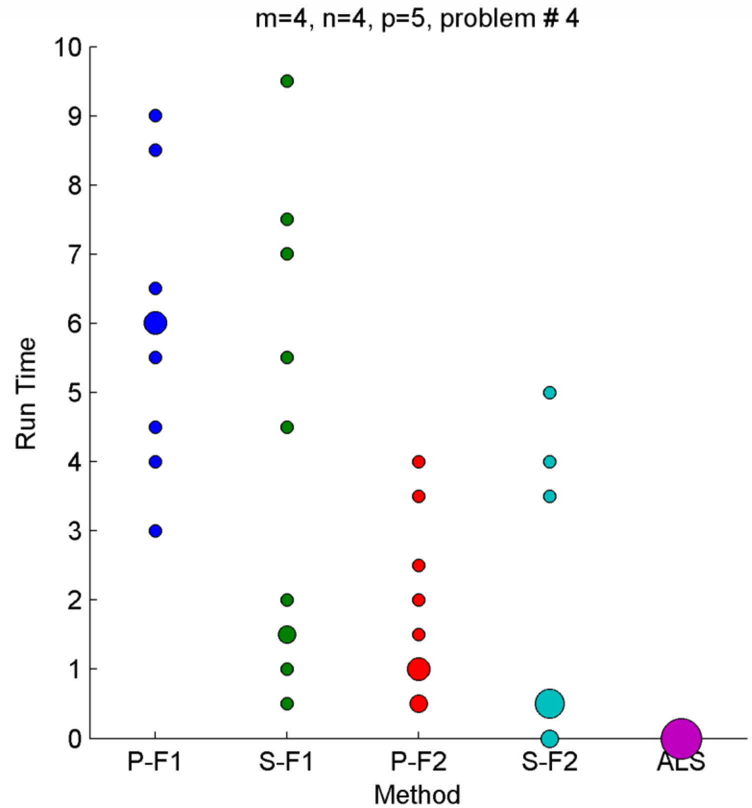




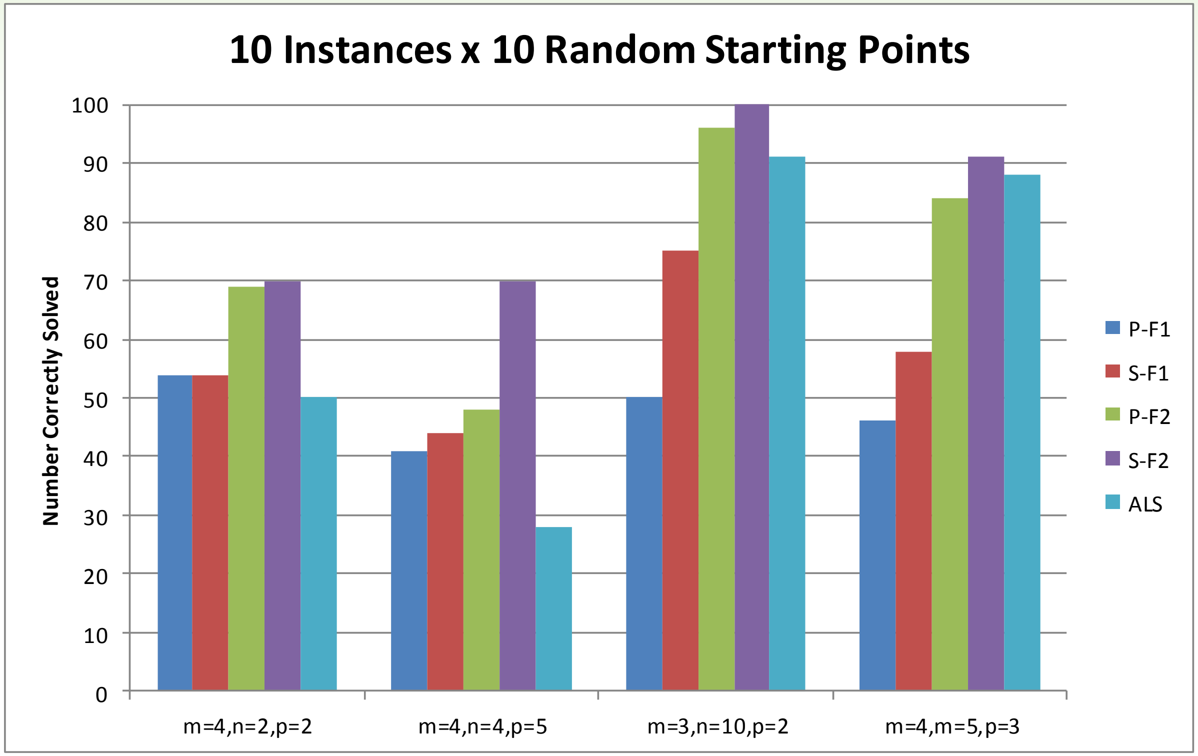
Poblano solves more problems for f_1

Sample Results: Results Vary





Overall Results



$$f_1(\boldsymbol{\lambda}, \mathbf{X}) = \sum_{i_1, i_2, \cdots, i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \left(\mathbf{x}_k^m \right)_{\mathbf{i}} \right)^2$$

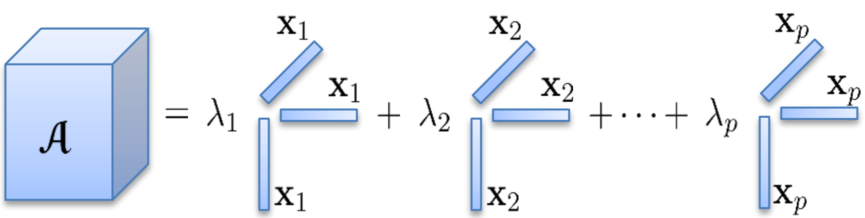
$$f_2(\boldsymbol{\lambda}, \mathbf{X}) = \sum_{i_1 \le i_2 \le \cdots \le i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \left(\mathbf{x}_k^m \right)_{\mathbf{i}} \right)^2$$

$$p(\mathbf{X}; \gamma) = \gamma \sum_{k=1}^{p} (\mathbf{x}_k^{\mathsf{T}} \mathbf{x}_k - 1)^2$$

- Formulation 2 better than Formulation 1
- SNOPT better than Poblano
- ALS can be competitive

Conclusions & Future Work

$$\mathcal{A} = \sum_{k=1}^{p} \lambda_k \, \mathbf{x}_k^m$$



$$f_1(\boldsymbol{\lambda}, \mathbf{X}) = \sum_{i_1, i_2, \cdots, i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \, (\mathbf{x}_k^m)_{\mathbf{i}} \right)^2$$

$$f_2(\boldsymbol{\lambda}, \mathbf{X}) = \sum_{i_1 \leq i_2 \leq \cdots \leq i_m} \left(a_{\mathbf{i}} - \sum_{k=1}^{p} \lambda_k \, (\mathbf{x}_k^m)_{\mathbf{i}} \right)^2$$

$$p(\mathbf{X}; \gamma) = \gamma \sum_{k=1}^{p} (\mathbf{x}_k^{\mathsf{T}} \mathbf{x}_k - 1)^2$$

- Summary
  - Nonconvex optimization problem
  - Formulation and solver impact solution
- Future work on symmetric decompositions
  - Another formulation as a generalized tensor eigenproblem
    - Allows SDP formulation
  - More extensive numerical results
  - Sensitivity to noise
  - Sensitivity to collinearity
  - Nonnegative factorizations
- Related work
  - Statistical methods for choosing rank