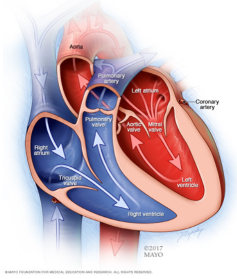
SIAM Conference on Applied Linear Algebra (SIAM-ALA18)

SUSTain: Scalable Unsupervised Scoring for Tensors and its Application to Phenotyping

Ioakeim Perros[1], Evangelos E. Papalexakis[2], Haesun Park[1], Richard Vuduc[1], Xiaowei Yan[3], Christopher deFilippi[4], Walter F. Stewart[3], Jimeng Sun[1]
Georgia Tech[1], UC Riverside[2], Sutter Health[3], Inova Heart and Vascular Institute[4]

To appear in the proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Clinical Phenotyping from Electronic Health Records



- Phenotype: set of measurable markers of a disease
  - e.g., what are the diagnoses and medications shared by various HF subtypes?

✗ Manual derivation of phenotypes is impractical (time-consuming chart reviewing)

✓ Goal: automatic derivation through EHR data

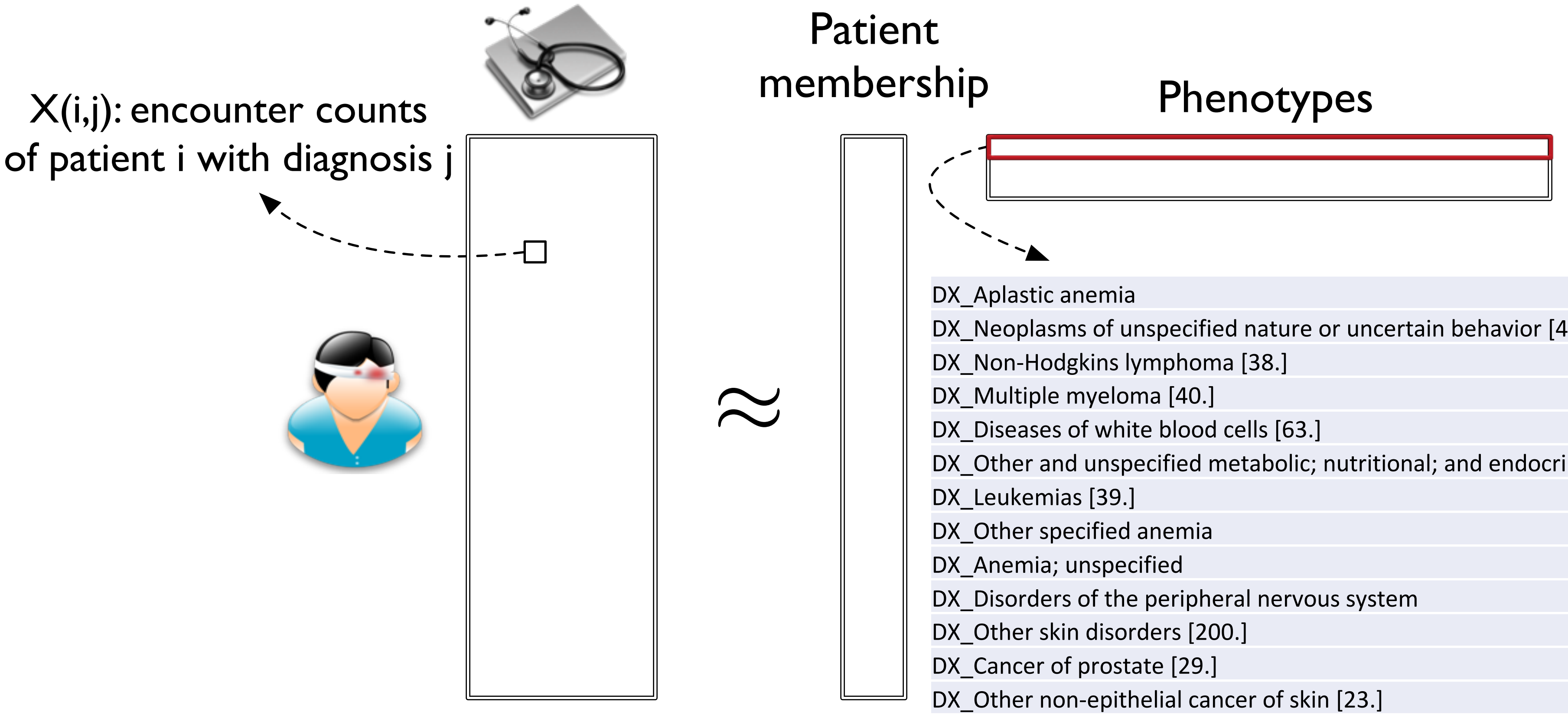
- No cases/controls labels assumption → targeting unsupervised techniques

Image source: https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-causes/syc-20373142

Richesson, Rachel L., et al. "Clinical phenotyping in selected national networks: Demonstrating the need for high-throughput, portable, and computational methods." Artificial intelligence in medicine 2016

Yadav, Pranjul, et al. "Mining Electronic Health Records: A Survey." arXiv 2017

Nonnegative Matrix/Tensor Factorization for Phenotyping



X(i,j): encounter counts of patient i with diagnosis j

Patient membership

Phenotypes

DX_Aplastic anemia	141.330727
DX_Neoplasms of unspecified nature or uncertain behavior [44.]	106.006925
DX_Non-Hodgkins lymphoma [38.]	83.34961
DX_Multiple myeloma [40.]	43.558884
DX_Diseases of white blood cells [63.]	28.197472
DX_Other and unspecified metabolic; nutritional; and endocrine disorders	27.291268
DX_Leukemias [39.]	26.831567
DX_Other specified anemia	19.721602
DX_Anemia; unspecified	17.325401
DX_Disorders of the peripheral nervous system	17.029972
DX_Other skin disorders [200.]	16.006847
DX_Cancer of prostate [29.]	12.103675
DX_Other non-epithelial cancer of skin [23.]	11.859002
DX_Nausea and vomiting [250.]	10.752094
DX_Phlebitis and thrombophlebitis	8.905566
DX_Allergic reactions [253.]	7.650406
DX_Other deficiency anemia	6.893146
DX_Other and unspecified lower respiratory disease	5.3113
DX_Iron deficiency anemia	4.957122
DX_Cardiomyopathy	4.917198
DX_Cataract [86.]	4.469955

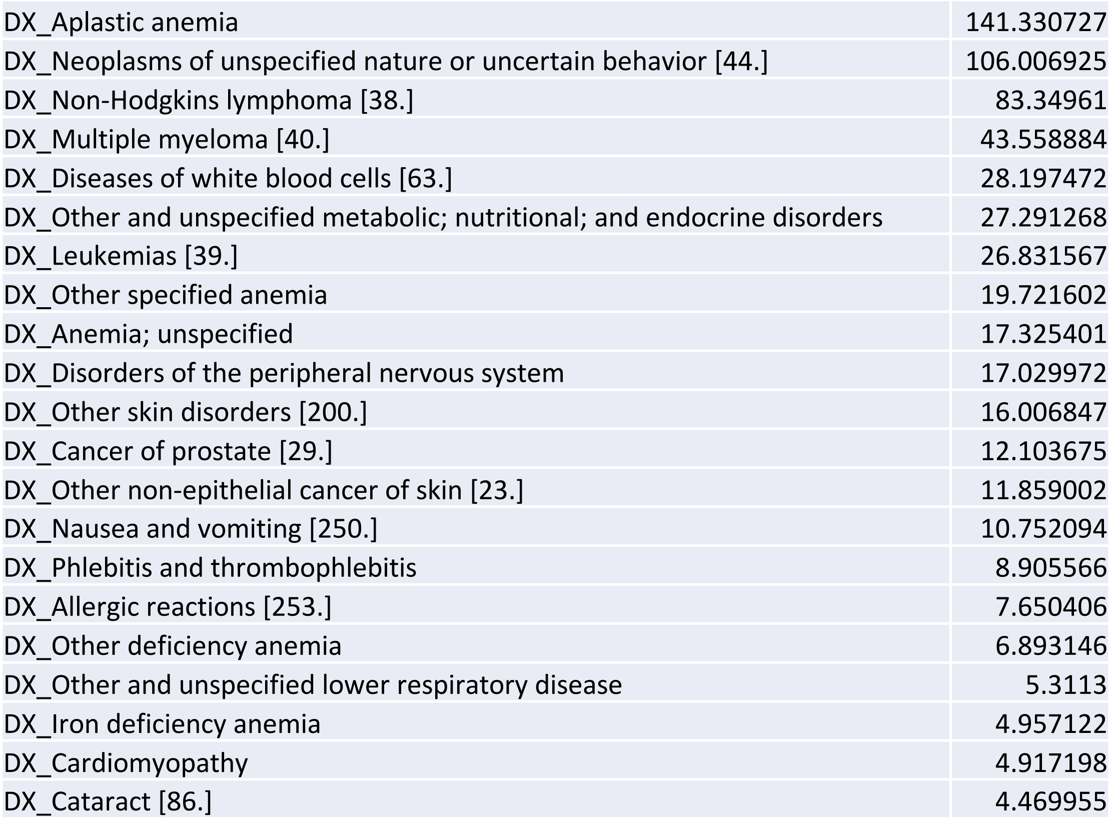
$$\min\left\{||\mathbf{X}-\mathbf{U}\mathbf{V}^T||_F^2 \;\middle|\; \mathbf{U}\geq 0, \mathbf{V}\geq 0\right\}$$

Nonnegative Matrix/Tensor Factorization for Phenotyping

Issues with representing integer data with real-valued factors

$$\min \{||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 \mid \mathbf{U} \geq 0, \mathbf{V} \geq 0\}$$

- Real factors are no longer interpretable as frequencies (input data)
- Arbitrary ranges and relative differences between elements
  - Hard to choose importance threshold
- Practitioners may be more familiar with scoring-based systems
  - e.g., medicine – risk scores



DX_Aplastic anemia	141.330727
DX_Neoplasms of unspecified nature or uncertain behavior [44.]	106.006925
DX_Non-Hodgkins lymphoma [38.]	83.34961
DX_Multiple myeloma [40.]	43.558884
DX_Diseases of white blood cells [63.]	28.197472
DX_Other and unspecified metabolic; nutritional; and endocrine disorders	27.291268
DX_Leukemias [39.]	26.831567
DX_Other specified anemia	19.721602
DX_Anemia; unspecified	17.325401
DX_Disorders of the peripheral nervous system	17.029972
DX_Other skin disorders [200.]	16.006847
DX_Cancer of prostate [29.]	12.103675
DX_Other non-epithelial cancer of skin [23.]	11.859002
DX_Nausea and vomiting [250.]	10.752094
DX_Phlebitis and thrombophlebitis	8.905566
DX_Allergic reactions [253.]	7.650406
DX_Other deficiency anemia	6.893146
DX_Other and unspecified lower respiratory disease	5.3113
DX_Iron deficiency anemia	4.957122
DX_Cardiomyopathy	4.917198
DX_Cataract [86.]	4.469955

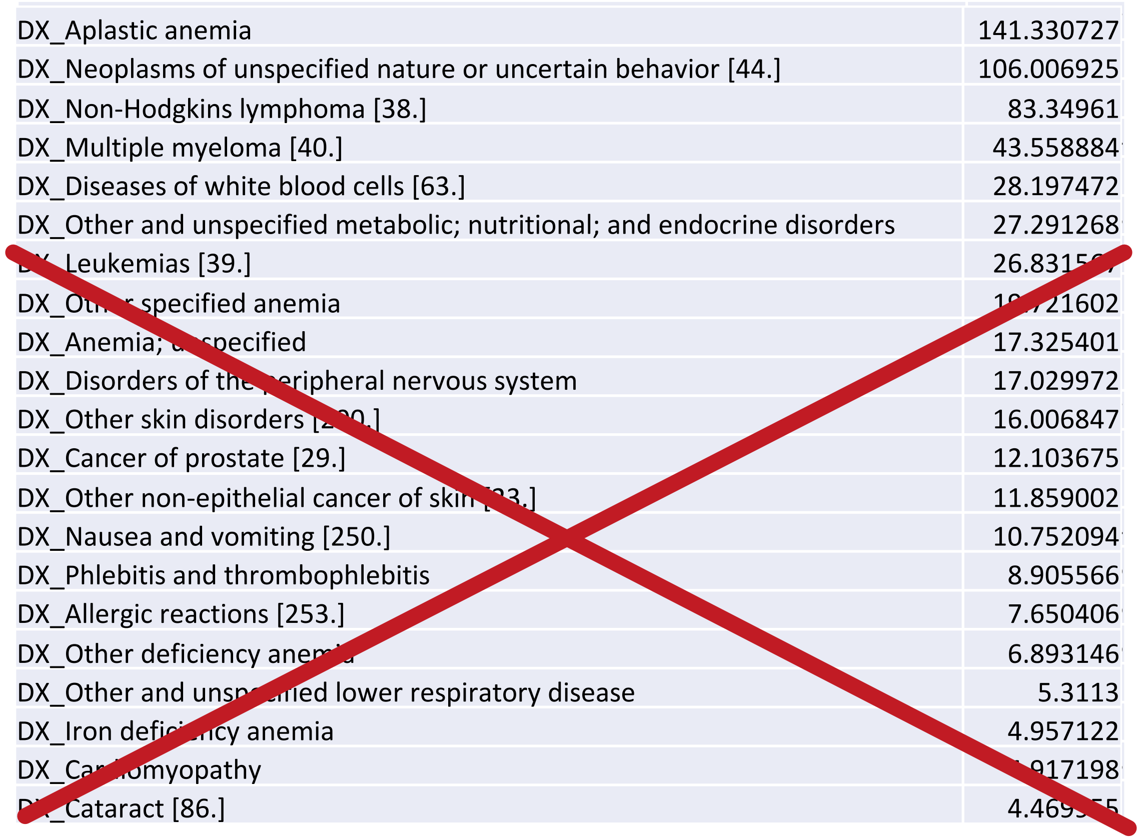
Nonnegative Matrix/Tensor Factorization for Phenotyping

Issues with representing integer data with real-valued factors

$$\min \{ ||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 \mid \mathbf{U} \geq 0, \mathbf{V} \geq 0 \}$$

- Ad-hoc heuristics are employed without formal justification

1. Arbitrary hard thresholding often leading to poor fit



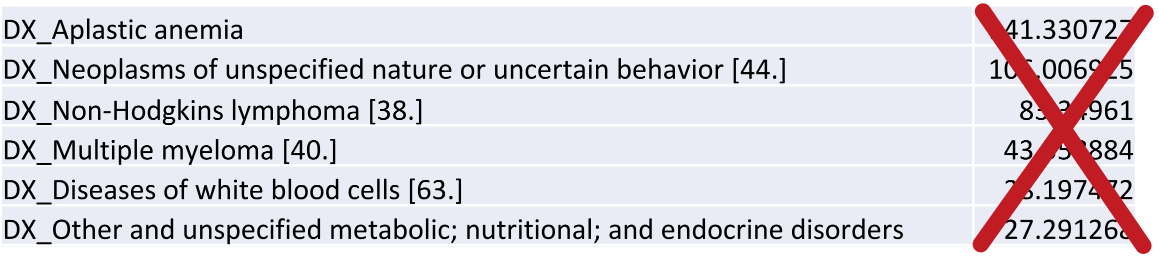
DX_Aplastic anemia	141.330727
DX_Neoplasms of unspecified nature or uncertain behavior [44.]	106.006925
DX_Non-Hodgkins lymphoma [38.]	83.34961
DX_Multiple myeloma [40.]	43.558884
DX_Diseases of white blood cells [63.]	28.197472
DX_Other and unspecified metabolic; nutritional; and endocrine disorders	27.291268
DX_Leukemias [39.]	26.8315[illegible]
DX_Other specified anemia	1[illegible].721602
DX_Anemia; unspecified	17.325401
DX_Disorders of the peripheral nervous system	17.029972
DX_Other skin disorders [2[illegible]0.]	16.006847
DX_Cancer of prostate [29.]	12.103675
DX_Other non-epithelial cancer of skin [[illegible]3.]	11.859002
DX_Nausea and vomiting [250.]	10.752094
DX_Phlebitis and thrombophlebitis	8.905566
DX_Allergic reactions [253.]	7.650406
DX_Other deficiency anemia	6.893146
DX_Other and unspecified lower respiratory disease	5.3113
DX_Iron deficiency anemia	4.957122
DX_Cardiomyopathy	[illegible].917198
DX_Cataract [86.]	4.469[illegible]55

Nonnegative Matrix/Tensor Factorization for Phenotyping

Issues with representing integer data with real-valued factors

$$\min \{ ||\mathbf{X} - \mathbf{U}\mathbf{V}^T||_F^2 \mid \mathbf{U} \geq 0, \mathbf{V} \geq 0 \}$$

- Ad-hoc heuristics are employed without formal justification



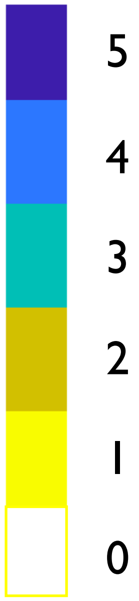
DX_Aplastic anemia	41.330727
DX_Neoplasms of unspecified nature or uncertain behavior [44.]	106.006925
DX_Non-Hodgkins lymphoma [38.]	83.334961
DX_Multiple myeloma [40.]	43.053884
DX_Diseases of white blood cells [63.]	23.197472
DX_Other and unspecified metabolic; nutritional; and endocrine disorders	27.291268

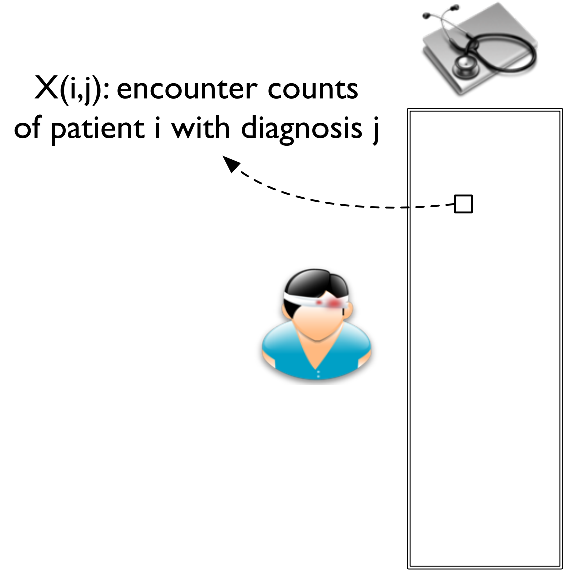
1. Arbitrary hard thresholding often leading to poor fit
2. Hidden values omitting potentially useful information
    - e.g., feature importance

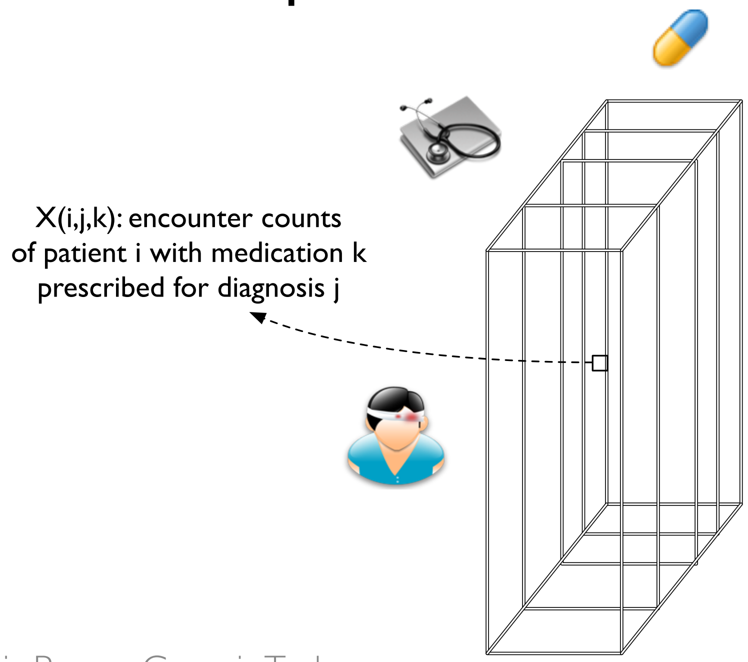
SUSTain: Scalable Unsupervised Scoring for Tensors

Overview (1)

- Integer-constrained factorization methodology
  - Factor matrices take scores constrained from a small integer set
  - Straightforward interpretation: distinct levels of feature importance
- Can handle both matrix and general tensor inputs



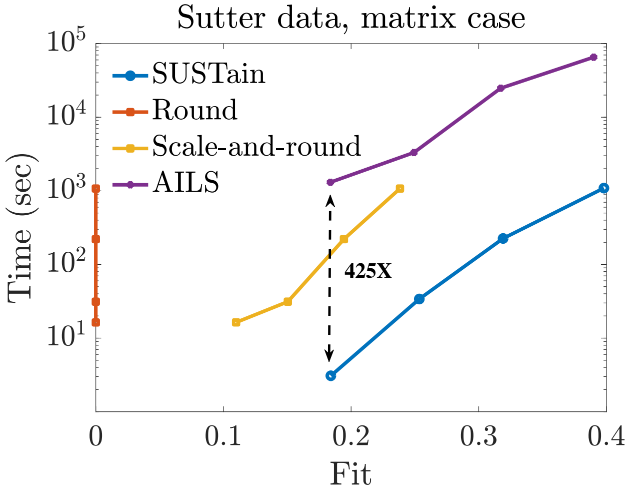

X(i,j): encounter counts of patient i with diagnosis j


X(i,j,k): encounter counts of patient i with medication k prescribed for diagnosis j

SUSTain: Scalable Unsupervised Scoring for Tensors

Overview (2)

- Problem partitioning → efficient and optimal solutions of integer-constraint subproblems
- Order of subproblems' solution → reuse of shared intermediate results
- Speedups up to 425X



- Case study on heart failure phenotyping
  - Cardiologist found 87% of phenotypes as clinically-meaningful

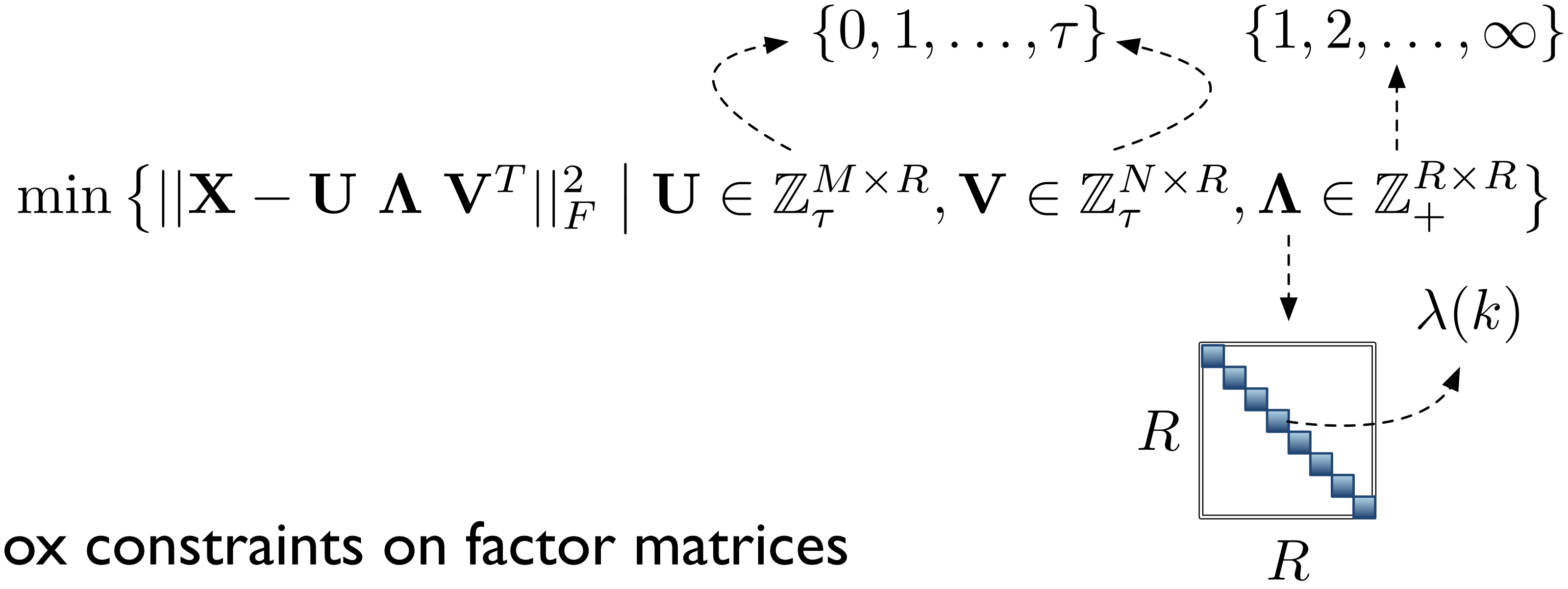
Hyperlipidemia	Score
Rx_HMG CoA Reductase Inhibitors	3
Dx_Disorders of lipid metabolism	1

Related Work

- Dong et al. → integer factorization based on integer least squares (Chang et al.)
  - Orders of magnitude slower achieving the same level of accuracy
- Kolda et al. → ternary ({-1, 0, 1}) factorization for compression purposes
  - Not easy to extend to general (nonnegative) integer box constraints
- Koyutürk et al. (among many others) → strictly binary factorization
  - Does not capture the quantity embedded in the input data which reveals important information

Dong, Bo, Matthew M. Lin, and Haesun Park. "Integer matrix approximation and data mining." Journal of scientific computing 75.1 (2018): 198-224.
Chang, Xiao-Wen, and Qing Han. "Solving box-constrained integer least squares problems." IEEE Transactions on wireless communications 7.1 (2008).
Kolda, Tamara G., and Dianne P. O'leary. "A semidiscrete matrix decomposition for latent semantic indexing information retrieval." ACM Transactions on Information Systems (TOIS)16.4 (1998): 322-346.
Koyutürk, Mehmet, Ananth Grama, and Naren Ramakrishnan. "Nonorthogonal decomposition of binary matrices for bounded-error data compression and analysis." ACM Transactions on Mathematical Software (TOMS) 32.1 (2006): 33-69.

SUSTain objective for matrix input (SUSTain$_M$)



$$\min\left\{||\mathbf{X} - \mathbf{U}\ \mathbf{\Lambda}\ \mathbf{V}^T||_F^2 \;\middle|\; \mathbf{U} \in \mathbb{Z}_\tau^{M\times R}, \mathbf{V} \in \mathbb{Z}_\tau^{N\times R}, \mathbf{\Lambda} \in \mathbb{Z}_+^{R\times R}\right\}$$

- Integer box constraints on factor matrices
- λ absorbs any scaling of each rank-1 component
  - Due to the integer constraint, λ cannot be obtained through normalization (as in NMF)
- $\mathbb{Z}_\tau$ can vary for different factor matrices and can be negative
  - Presentation aligns with phenotyping application needs

SUSTain$_M$ fitting algorithm (1)

$$\min\{||\underbrace{\mathbf{X} - \sum_{r=1, r\neq k}^{R} \lambda(r)\ \mathbf{U}(:,r)\ \mathbf{V}(:,r)^T}_{\mathbf{R}_k} - \enclose{circle}{\lambda(k)}\ \mathbf{U}(:,k)\ \mathbf{V}(:,k)^T||_F^2$$

$$|\ \mathbf{U} \in \mathbb{Z}_\tau^{M\times R}, \mathbf{V} \in \mathbb{Z}_\tau^{N\times R}, \mathbf{\Lambda} \in \mathbb{Z}_+^{R\times R}\}$$

- Solving for each k-th rank-1 component separately (intuition behind HALS, Cichocki et al.)
- $\lambda(k)$ solution: second-order scalar equation

$$\lambda(k) \leftarrow max\left(1, round\left(\lambda(k) + \frac{\mathbf{V}(:,k)^T\ \left([\mathbf{X}^T\ \mathbf{U}]_{:,k}\ -\ \mathbf{V}\ \mathbf{\Lambda}\ [\mathbf{U}^T\ \mathbf{U}]_{:,k}\right)}{[\mathbf{U}^T\ \mathbf{U}]_{k,k}\ [\mathbf{V}^T\ \mathbf{V}]_{k,k}}\right)\right)$$

Cichocki, Andrzej, and Anh-Huy Phan. "Fast local algorithms for large scale nonnegative matrix and tensor factorizations." IEICE transactions on fundamentals of electronics, communications and computer sciences 92.3 (2009): 708-721.

SUSTain$_M$ fitting algorithm (2)

$$\min\{||\underbrace{\mathbf{X} - \sum_{r=1,r\neq k}^{R} \lambda(r)\ \mathbf{U}(:,r)\ \mathbf{V}(:,r)^T}_{\mathbf{R}_k} - \lambda(k)\ \mathbf{U}(:,k)\ \mathbf{V}(:,k)^T||_F^2$$

$$|\ \mathbf{U} \in \mathbb{Z}_\tau^{M\times R}, \mathbf{V} \in \mathbb{Z}_\tau^{N\times R}, \mathbf{\Lambda} \in \mathbb{Z}_+^{R\times R}\}$$

- $\boldsymbol{V}(:,k)$ solution: optimal scaling lemma (Bro and Sidiropoulos, 1998)

$$\min\left\{||\mathbf{Y} - \mathbf{x}\ \mathbf{b}^T||_2^2 \ \middle|\ \mathbf{b} \in C\right\} = \Pi_C(\beta) \qquad \beta = \frac{\mathbf{x}^T\ \mathbf{Y}}{\mathbf{x}^T\ \mathbf{x}}$$

- Optimal solution of constrained problem: projection of unconstrained solution to C

$$\mathbf{b} \leftarrow \mathbf{V}(:,k) + \frac{[\mathbf{X}^T\ \mathbf{U}]_{:,k} - \mathbf{V}\ \mathbf{\Lambda}\ [\mathbf{U}^T\ \mathbf{U}]_{:,k}}{[\mathbf{U}^T\ \mathbf{U}]_{k,k}\ \lambda(k)} \qquad \mathbf{V}(:,k) \leftarrow min\left(max\left(round\left(\mathbf{b}\right),0\right),\tau\right)$$

Bro, Rasmus, and Nicholaos D. Sidiropoulos. "Least squares algorithms under unimodality and non-negativity constraints." Journal of Chemometrics 12.4 (1998): 223-247.

SUSTain$_M$ fitting algorithm (3)

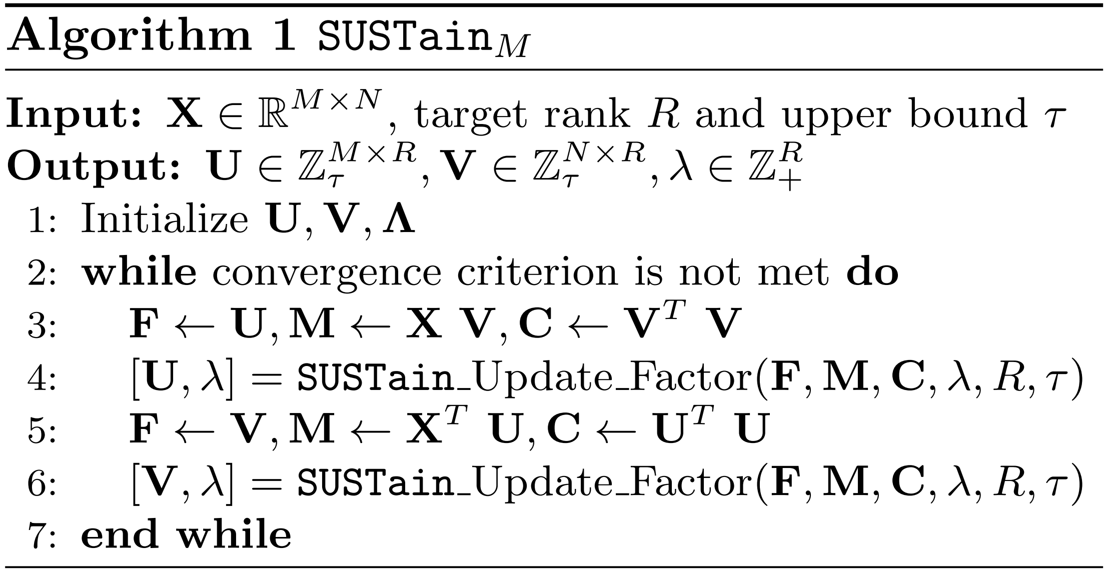
$$\lambda(k) \leftarrow max\left(1, round\left(\lambda(k) + \frac{\mathbf{V}(:,k)^T\left([\mathbf{X}^T\ \mathbf{U}]_{:,k}\ -\ \mathbf{V}\ \boldsymbol{\Lambda}\ [\mathbf{U}^T\ \mathbf{U}]_{:,k}\right)}{[\mathbf{U}^T\ \mathbf{U}]_{k,k}\ [\mathbf{V}^T\ \mathbf{V}]_{k,k}}\right)\right)$$

$$\mathbf{b} \leftarrow \mathbf{V}(:,k) + \frac{[\mathbf{X}^T\ \mathbf{U}]_{:,k} - \mathbf{V}\ \boldsymbol{\Lambda}\ [\mathbf{U}^T\ \mathbf{U}]_{:,k}}{[\mathbf{U}^T\ \mathbf{U}]_{k,k}\ \lambda(k)}$$

$$\mathbf{V}(:,k) \leftarrow min\left(max\left(round\left(\mathbf{b}\right), 0\right), \tau\right)$$

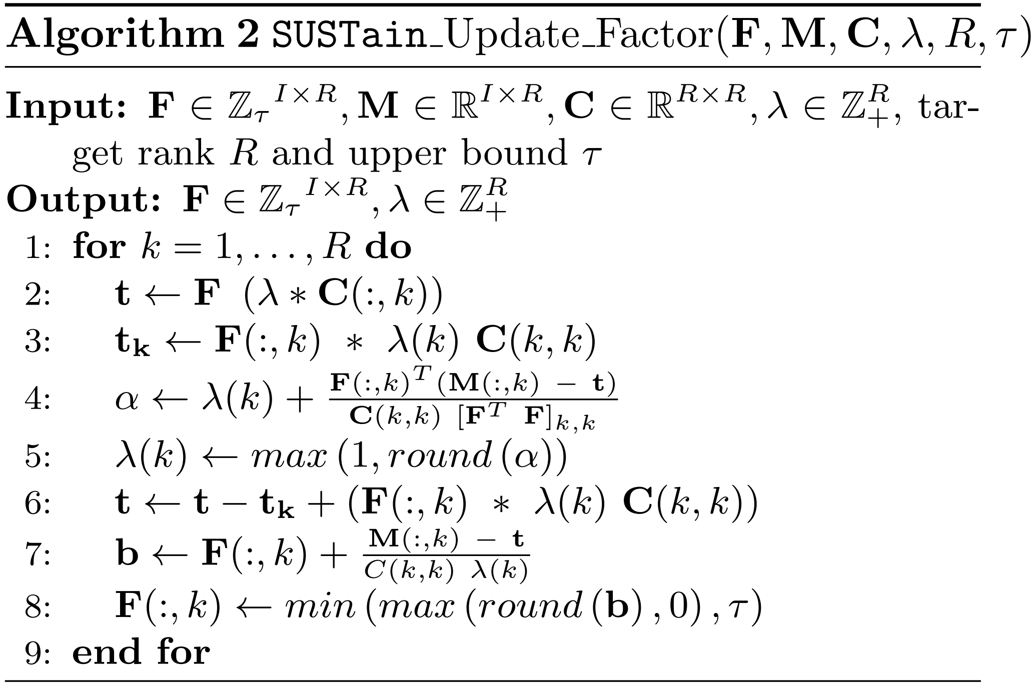
- Identify computationally expensive shared intermediate results → update $\lambda(k)$ and $\boldsymbol{V}(:,k)$ during the same iteration
- Only need to re-compute $\mathbf{t} := \mathbf{V}\ \boldsymbol{\Lambda}\ [\mathbf{U}^T\ \mathbf{U}]_{:,k}$ after having updated $\lambda(k)$
- Symmetric update for $\boldsymbol{U}(:,k)$

SUSTain$_M$ fitting algorithm (4)



Algorithm 1 `SUSTain`$_M$

```
Input: $\mathbf{X} \in \mathbb{R}^{M \times N}$, target rank $R$ and upper bound $\tau$
Output: $\mathbf{U} \in \mathbb{Z}_\tau^{M \times R}, \mathbf{V} \in \mathbb{Z}_\tau^{N \times R}, \lambda \in \mathbb{Z}_+^R$
Initialize $\mathbf{U}, \mathbf{V}, \mathbf{\Lambda}$
while convergence criterion is not met do
    $\mathbf{F} \leftarrow \mathbf{U}, \mathbf{M} \leftarrow \mathbf{X}\ \mathbf{V}, \mathbf{C} \leftarrow \mathbf{V}^T\ \mathbf{V}$
    $[\mathbf{U}, \lambda] =$ SUSTain_Update_Factor$(\mathbf{F}, \mathbf{M}, \mathbf{C}, \lambda, R, \tau)$
    $\mathbf{F} \leftarrow \mathbf{V}, \mathbf{M} \leftarrow \mathbf{X}^T\ \mathbf{U}, \mathbf{C} \leftarrow \mathbf{U}^T\ \mathbf{U}$
    $[\mathbf{V}, \lambda] =$ SUSTain_Update_Factor$(\mathbf{F}, \mathbf{M}, \mathbf{C}, \lambda, R, \tau)$
end while
```



Algorithm 2 `SUSTain`_Update_Factor$(\mathbf{F}, \mathbf{M}, \mathbf{C}, \lambda, R, \tau)$

```
Input: $\mathbf{F} \in \mathbb{Z}_\tau{}^{I \times R}, \mathbf{M} \in \mathbb{R}^{I \times R}, \mathbf{C} \in \mathbb{R}^{R \times R}, \lambda \in \mathbb{Z}_+^R$, target rank $R$ and upper bound $\tau$
Output: $\mathbf{F} \in \mathbb{Z}_\tau{}^{I \times R}, \lambda \in \mathbb{Z}_+^R$
for $k = 1, \ldots, R$ do
    $\mathbf{t} \leftarrow \mathbf{F}\ (\lambda * \mathbf{C}(:,k))$
    $\mathbf{t_k} \leftarrow \mathbf{F}(:,k)\ *\ \lambda(k)\ \mathbf{C}(k,k)$
    $\alpha \leftarrow \lambda(k) + \frac{\mathbf{F}(:,k)^T (\mathbf{M}(:,k)\ -\ \mathbf{t})}{\mathbf{C}(k,k)\ [\mathbf{F}^T\ \mathbf{F}]_{k,k}}$
    $\lambda(k) \leftarrow max\,(1, round\,(\alpha))$
    $\mathbf{t} \leftarrow \mathbf{t} - \mathbf{t_k} + (\mathbf{F}(:,k)\ *\ \lambda(k)\ \mathbf{C}(k,k))$
    $\mathbf{b} \leftarrow \mathbf{F}(:,k) + \frac{\mathbf{M}(:,k)\ -\ \mathbf{t}}{C(k,k)\ \lambda(k)}$
    $\mathbf{F}(:,k) \leftarrow min\,(max\,(round\,(\mathbf{b})\,, 0)\,, \tau)$
end for
```

Extension for tensor input

$$\min\{||\mathcal{X} - \sum_{r=1}^{R} \lambda(r)\ \mathbf{A}^{(1)}(:,r)\ \circ \ldots\ \circ\ \mathbf{A}^{(d)}(:,r)||_F^2$$

$$|\ \mathbf{A}^{(n)} \in \mathbb{Z}_\tau^{I_n \times R}, \lambda(r) \in \mathbb{Z}_+\}$$

- Constrained version of CP model (Hitchcock, Harshman, Carroll and Chang)

$$\mathcal{R}_k := \mathcal{X} - \sum_{r=1, r\neq k}^{R} \lambda(r)\ \mathbf{A}^{(1)}(:,r)\ \circ \ldots\ \circ\ \mathbf{A}^{(d)}(:,r)$$

$$\min\{||\mathcal{R}_k - \lambda(k)\ \mathbf{A}^{(1)}(:,k) \circ \ldots \circ \mathbf{A}^{(d)}(:,k)||_F^2 |\ \mathbf{A}^{(n)} \in \mathbb{Z}_\tau^{I_n \times R}, \lambda(k) \in \mathbb{Z}_+\}$$

Hitchcock, Frank L. "The expression of a tensor or a polyadic as a sum of products." Studies in Applied Mathematics 6.1-4 (1927): 164-189.
Harshman, Richard A. "Foundations of the parafac procedure: models and conditions for an" explanatory" multimodal factor analysis." (1970): 84.
Carroll, J. Douglas, and Jih-Jie Chang. "Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition." Psychometrika 35.3 (1970): 283-319.

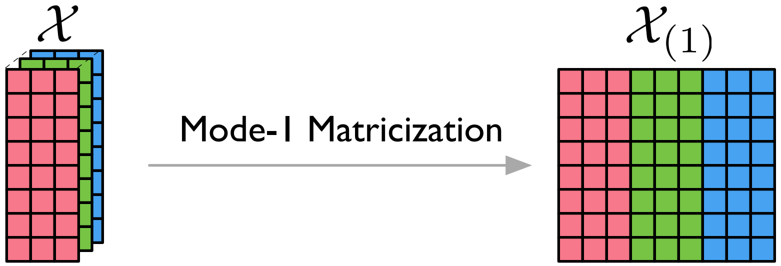
MTTKRP

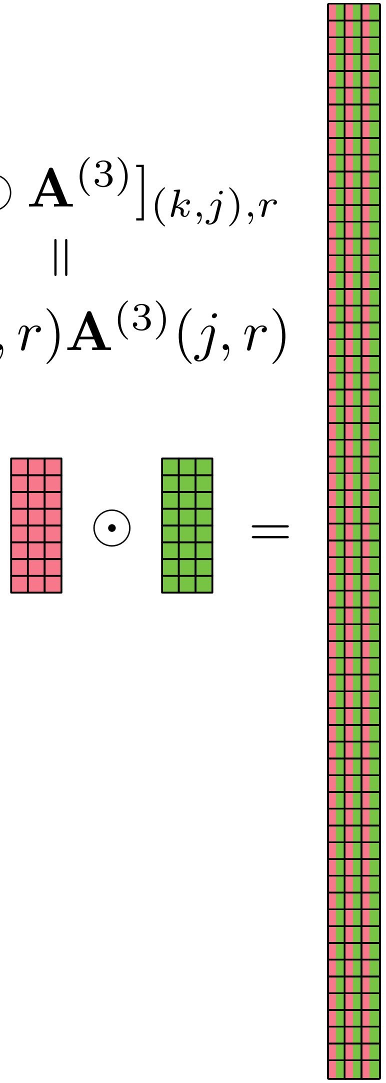
Matricized Tensor Times Khatri-Rao Product

$$[\mathbf{A}^{(2)} \odot \mathbf{A}^{(3)}]_{(k,j),r}$$

$$\|$$

$$\mathbf{A}^{(2)}(k,r)\mathbf{A}^{(3)}(j,r)$$

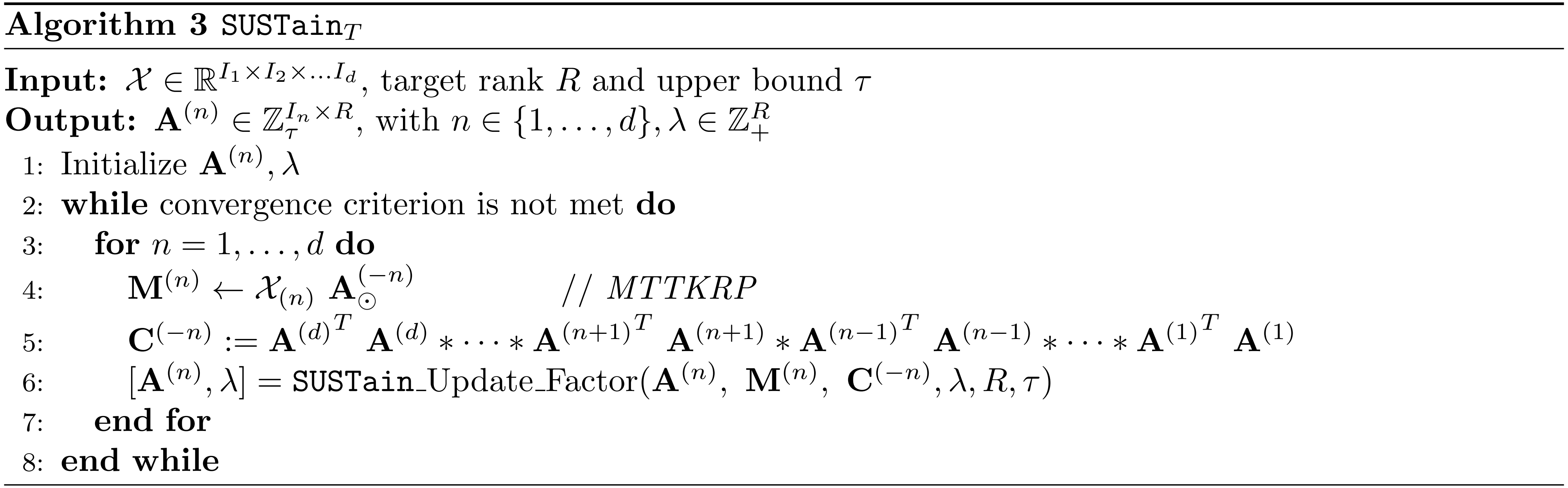




$$\text{Mode-1 MTTKRP: } \mathcal{X}_{(1)}\mathbf{A}_{\odot}^{(-1)}$$

Bader, Brett W., and Tamara G. Kolda. "Efficient MATLAB computations with sparse and factored tensors." SIAM Journal on Scientific Computing 30.1 (2007): 205-231.

SUSTain$_T$ fitting algorithm

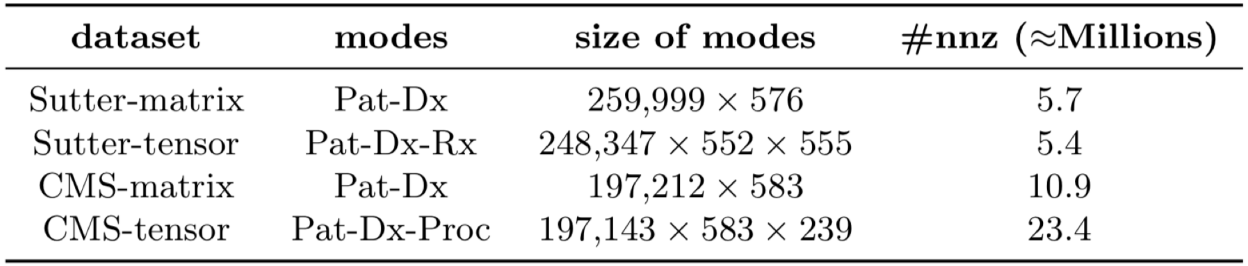


Algorithm 3 SUSTain$_T$

```
Input: $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \dots I_d}$, target rank $R$ and upper bound $\tau$
Output: $\mathbf{A}^{(n)} \in \mathbb{Z}_{\tau}^{I_n \times R}$, with $n \in \{1, \dots, d\}, \lambda \in \mathbb{Z}_{+}^{R}$
Initialize $\mathbf{A}^{(n)}, \lambda$
while convergence criterion is not met do
    for $n = 1, \dots, d$ do
        $\mathbf{M}^{(n)} \leftarrow \mathcal{X}_{(n)} \ \mathbf{A}_{\odot}^{(-n)}$        // MTTKRP
        $\mathbf{C}^{(-n)} := \mathbf{A}^{(d)^T} \ \mathbf{A}^{(d)} * \dots * \mathbf{A}^{(n+1)^T} \ \mathbf{A}^{(n+1)} * \mathbf{A}^{(n-1)^T} \ \mathbf{A}^{(n-1)} * \dots * \mathbf{A}^{(1)^T} \ \mathbf{A}^{(1)}$
        $[\mathbf{A}^{(n)}, \lambda] =$ SUSTain_Update_Factor($\mathbf{A}^{(n)}, \ \mathbf{M}^{(n)}, \ \mathbf{C}^{(-n)}, \lambda, R, \tau$)
    end for
end while
```

- Routine re-use from matrix case
- Can directly exploit already-developed scalable software for bottleneck MTTKRP (e.g., Bader and Kolda)

Data Description



dataset	modes	size of modes	#nnz (≈Millions)
Sutter-matrix	Pat-Dx	259,999 × 576	5.7
Sutter-tensor	Pat-Dx-Rx	248,347 × 552 × 555	5.4
CMS-matrix	Pat-Dx	197,212 × 583	10.9
CMS-tensor	Pat-Dx-Proc	197,143 × 583 × 239	23.4

- Sutter Palo Alto Medical Foundation Clinics
  - HF study (cases & controls)
  - ICD-9 dx codes → CCS level 4
  - Drugs represented through their therapeutic subclasses (ATCCS)
- CMS: publicly-available synthetic Medicare data (carrier claims, samples 1 & 2)
  - ICD-9 dx codes → CCS level 4
  - CPT procedure codes → CCS flat code grouper
  - Consider more data samples for scalability experiments

https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/DE_Syn_PUF.html

Baselines

- Round: rounds solutions of NMF/CP models to the nearest integer
- Scale-and-round: first scale the factor matrices before performing the rounding → partially alleviates zeroing out values less than 0.5
- AILS: Alternating Integer Least-Squares (ILS)
    - Extraction of λ through: $\min \left\{ ||(\boldsymbol{V} \odot \boldsymbol{U})\boldsymbol{\lambda} - vec(\boldsymbol{X})||_F^2 \;\middle|\; \boldsymbol{\lambda} \in \mathbb{Z}_+^R \right\}$
    - Non-scalable for tensor case: requires KRP of all factor matrices, failed even for the smallest target rank

Dong, Bo, Matthew M. Lin, and Haesun Park. "Integer matrix approximation and data mining." Journal of scientific computing 75.1 (2018): 198-224.
Chang, Xiao-Wen, and Qing Han. "Solving box-constrained integer least squares problems." IEEE Transactions on wireless communications 7.1 (2008).

Implementation & Evaluation

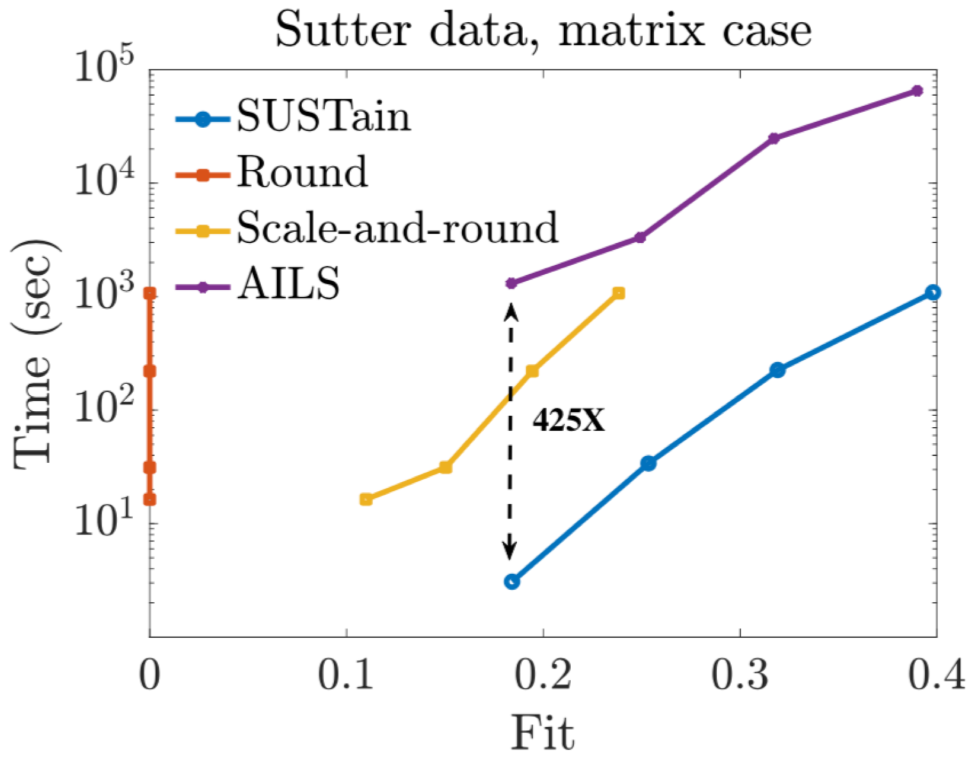
- MatlabR2017b: Tensor Toolbox, Nonnegfac-Matlab toolbox, MILES software for ILS
- Fit: $1 - ||\mathbf{X} - \hat{\mathbf{X}}||_F / ||\mathbf{X}||_F$
- Several initialization schemes (round, scale-and-round, random, random & sampling) for accuracy-time trade-off of SUSTain and AILS → pick solution providing the highest fit

B. W. Bader, T. G. Kolda, et al. Matlab tensor toolbox version 2.6. Available online, February 2015.
J. Kim, Y. He, and H. Park. Algorithms for nonnegative matrix and tensor factorizations: A unified view based on block coordinate descent framework. Journal of Global Optimization, 58(2):285–319, Feb. 2014.
X.-W. Chang and T. Zhou. Miles: Matlab package for solving mixed integer least squares problems. GPS Solutions, 11(4):289–294, 2007. Last updated: June 2016.

Fit-time trade-off: matrix case

R = {5, 10, 20, 40}

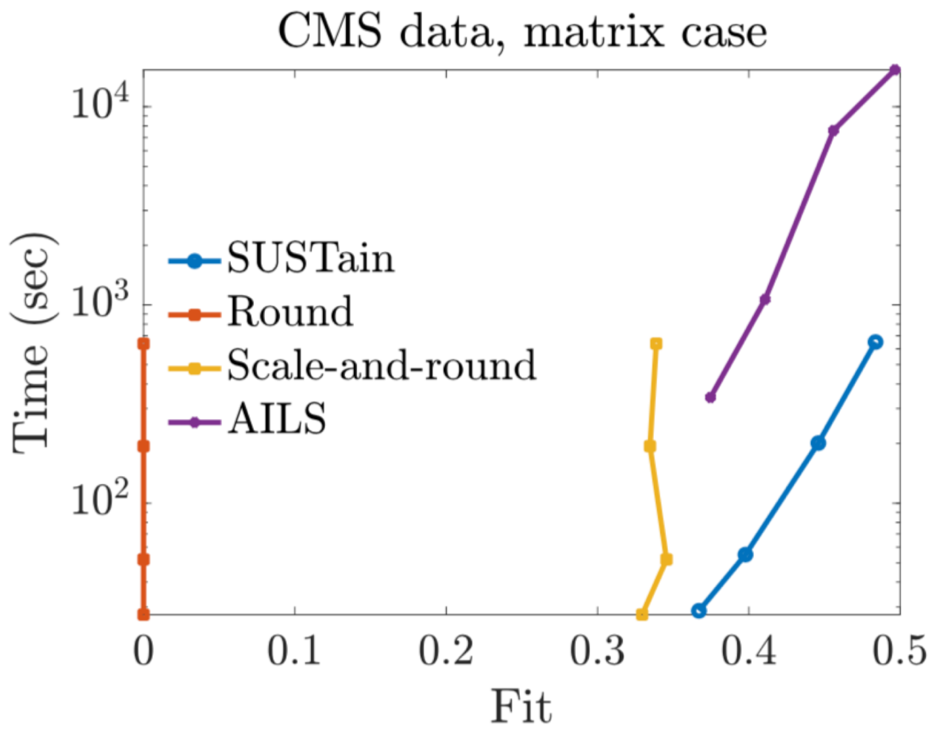
- Up to **425X faster** for the same fit against the most accurate baseline (3 seconds vs 22 mins)
- Orders of magnitude faster even for larger ranks (110X faster for R = 20)
- Up to 16% higher fit than scale-and-round heuristic



Fit-time trade-off: matrix case

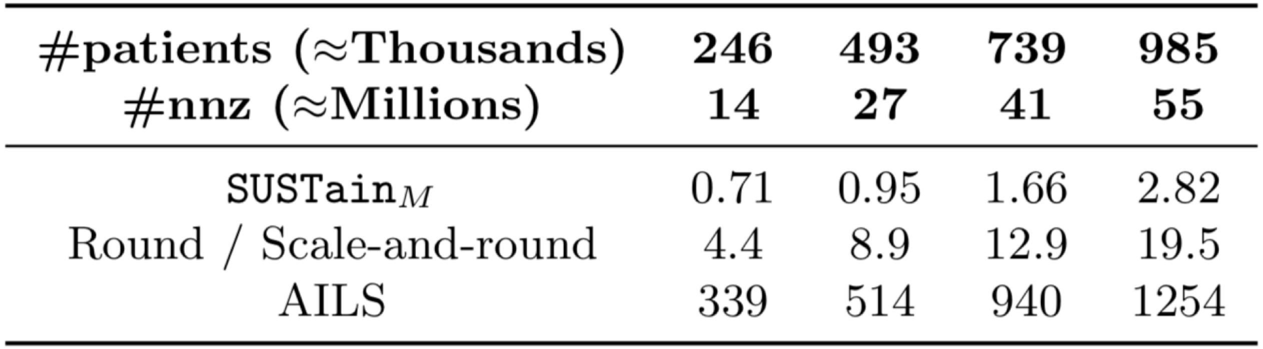
R = {5, 10, 20, 40}

- At least an order of magnitude speedup than AILS
- Up to 38X faster for R=20
- Up to 14% higher fit than scale-and-round



Scalability for increasingly larger #patients

Execution time of a single iteration

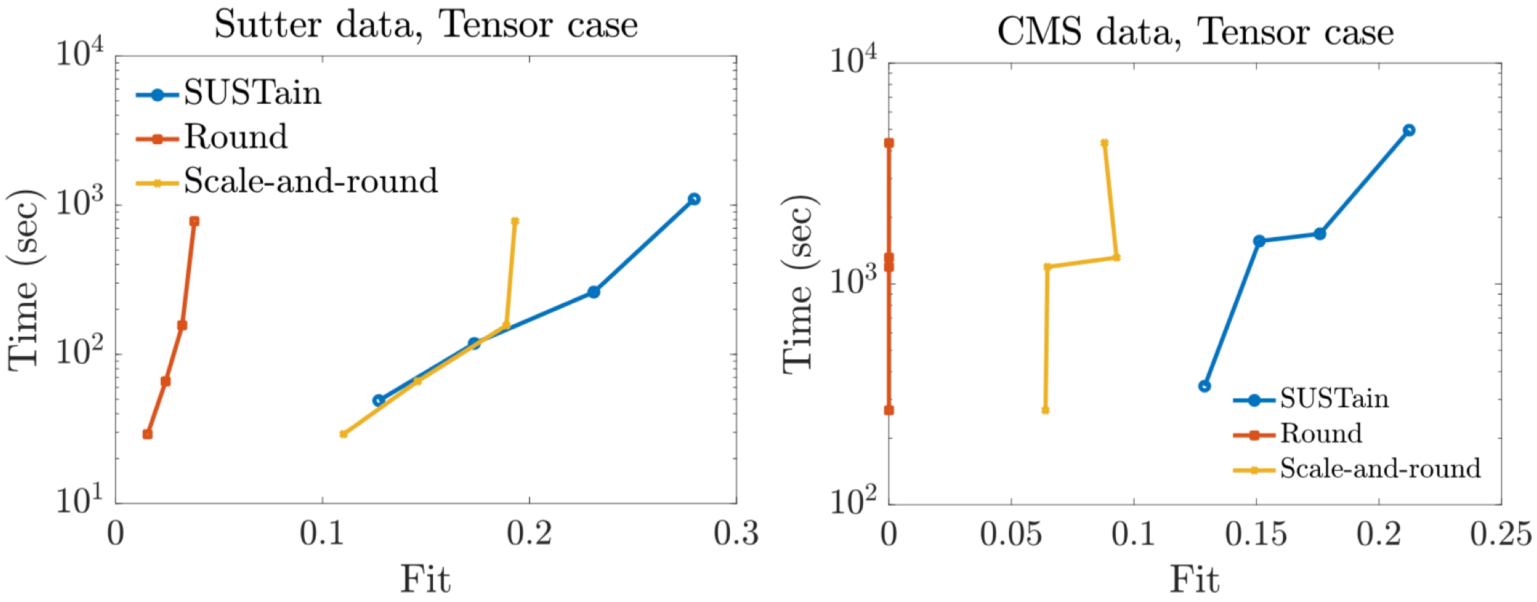


#patients (≈Thousands) **#nnz (≈Millions)**	**246** **14**	**493** **27**	**739** **41**	**985** **55**
$\texttt{SUSTain}_M$	0.71	0.95	1.66	2.82
Round / Scale-and-round	4.4	8.9	12.9	19.5
AILS	339	514	940	1254

- Round / Scale-and-round corresponds to NMF execution time
- Even for 985K patients, SUSTain executes very fast (3 seconds)

Fit-time trade-off: tensor case

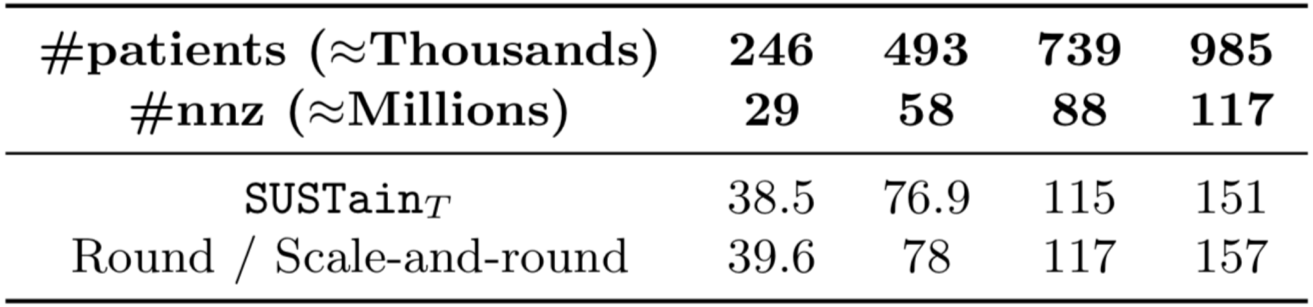
R = {5, 10, 20, 40}



- Up to 9% and 12% higher fit than heuristics
- Fit of scale-and-round decreases from R=20 to R=40 for CMS dataset
  - Heuristics may not fully exploit the available target rank

Scalability for increasingly larger #patients

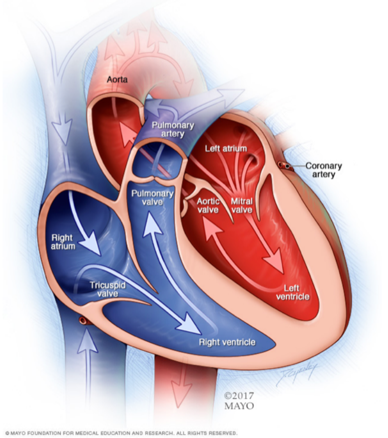
Execution time of a single iteration



#patients (≈Thousands)	**246**	**493**	**739**	**985**
#nnz (≈Millions)	**29**	**58**	**88**	**117**
SUSTain$_T$	38.5	76.9	115	151
Round / Scale-and-round	39.6	78	117	157

- Round / Scale-and-round corresponds to CP-ALS execution time
- SUSTain achieves linear scale-up w.r.t. increasing #patients
- Dominant cost is MTTKRP in both methods → comparable running times

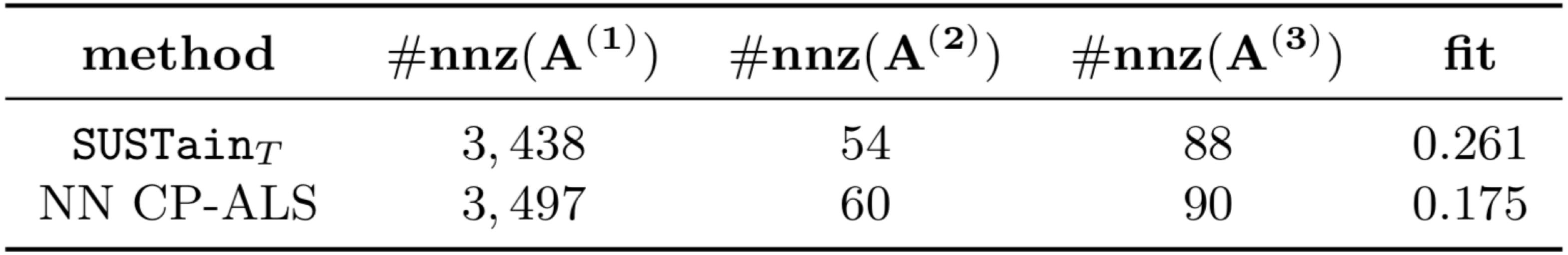
Case study: phenotyping HF patients



- CVD: leading cause of death worldwide
  - HF: dominant cause of morbidity and mortality
  - Recent evidence suggests heterogeneity in HF
- Case patients: (-1 year before, +1 year after) HF dx date
- 3,497 X 396 X 367 pat-dx-med tensor: 92,662 non-zeros
- #phenotypes choice: adaptation of work from Wu et al. for tensors
  - Promoting a target rank for which several runs with different initial points return reproducible factors
  - 15 phenotypes extracted

Image source: https://www.mayoclinic.org/diseases-conditions/heart-failure/symptoms-causes/syc-20373142
Stats source: World Health Organization (http://www.who.int/cardiovascular_diseases/en/)
S. Wu, A. Joseph, A. S. Hammonds, S. E. Celniker, B. Yu, and E. Frise. Stability-driven nonnegative matrix factorization to interpret spatial gene expression and build local gene networks. Proceedings of the National Academy of Sciences, 113(16):4290–4295, 2016.

SUSTain is concise and accurate

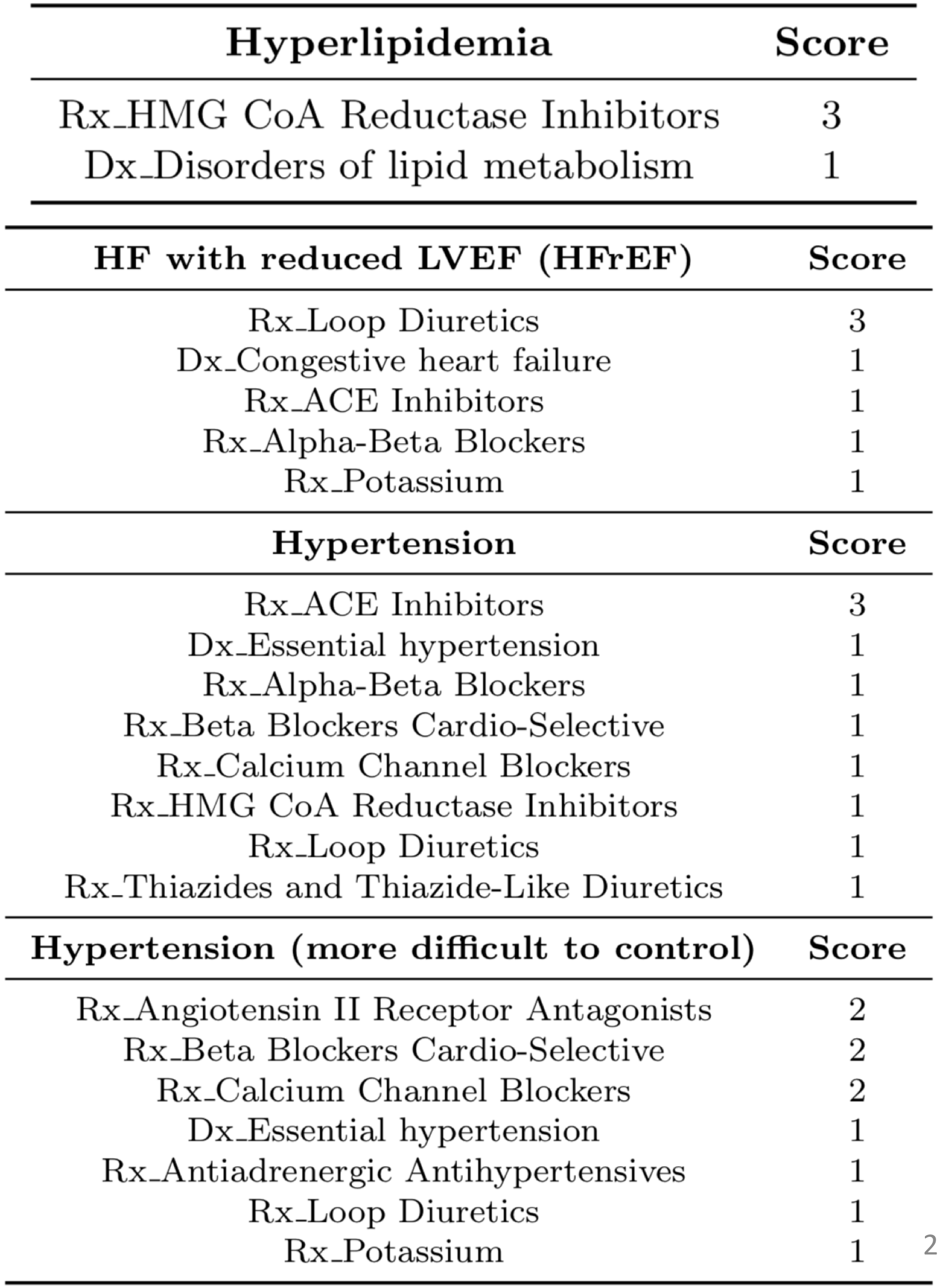


method	#nnz($\mathbf{A}^{(1)}$)	#nnz($\mathbf{A}^{(2)}$)	#nnz($\mathbf{A}^{(3)}$)	fit
SUSTain$_T$	$3,438$	54	88	0.261
NN CP-ALS	$3,497$	60	90	0.175

- SUSTain implicitly achieves sparsity → are the factors concise enough?
- Compare with NN CP-ALS model truncated to achieve the same level of sparsity
  - For the feature factors, consider top-k of each column (most important features per phenotype)
  - For the patient factors, consider top-k elements of each row (most important phenotypes per patient)
  - As would be done by a practitioner
- SUSTain achieves 8.6% increase in fit for the same level of sparsity

Phenotype discovery

- Other phenotype annotations provided by the cardiologist:
  - Persistent and chronic atrial fibrillation
  - Depression
  - Diabetes
  - Comorbidities of aging
  - Prior pulmonary embolism
- Overall, 13 out of 15 phenotype candidates were annotated as clinically meaningful



Hyperlipidemia	**Score**
Rx_HMG CoA Reductase Inhibitors	3
Dx_Disorders of lipid metabolism	1

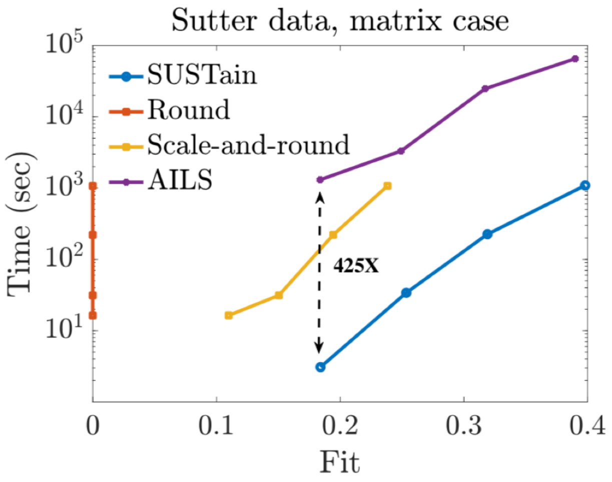
HF with reduced LVEF (HFrEF)	**Score**
Rx_Loop Diuretics	3
Dx_Congestive heart failure	1
Rx_ACE Inhibitors	1
Rx_Alpha-Beta Blockers	1
Rx_Potassium	1

Hypertension	**Score**
Rx_ACE Inhibitors	3
Dx_Essential hypertension	1
Rx_Alpha-Beta Blockers	1
Rx_Beta Blockers Cardio-Selective	1
Rx_Calcium Channel Blockers	1
Rx_HMG CoA Reductase Inhibitors	1
Rx_Loop Diuretics	1
Rx_Thiazides and Thiazide-Like Diuretics	1

Hypertension (more difficult to control)	**Score**
Rx_Angiotensin II Receptor Antagonists	2
Rx_Beta Blockers Cardio-Selective	2
Rx_Calcium Channel Blockers	2
Dx_Essential hypertension	1
Rx_Antiadrenergic Antihypertensives	1
Rx_Loop Diuretics	1
Rx_Potassium	1

Take-away

www.cc.gatech.edu/~iperros3/
perros@gatech.edu
Paper pre-print: goo.gl/s8yjxc



- Rounding of real-valued solutions does not always work
- Careful sub-problem partitioning leads to optimal and efficient solutions
- Order of updates defined to re-use shared intermediate results
- Overall, SUSTain outperforms several baselines
  - Either better fit or orders-of-magnitude speedup at a comparable fit
- 87% of phenotypes annotated as clinically meaningful
- Future work: factorizing ordinal values, establishing convergence results