Sparse Interactions: Identifying High-Dimensional Multilinear Systems via Compressed Sensing

Bobak Nazer and Robert D. Nowak
University of Wisconsin, Madison

Allerton 10/01/10
Motivation: Virus-Host Interaction

from Paul Ahlquist and Yoshihiro Kawaoka’s Labs at UW - Madison.
Motivation: Virus-Host Interaction

Add dsRNA of the *Drosophila* RNAi library (targeting to 13,071 *Drosophila* genes) to each well of 384-well microplates

Add DL1 cells to the plates

Infect with FVG-R virus

Measure *Renilla* luciferase activity to assess the efficiency of virus replication

from Paul Ahlquist and Yoshihiro Kawaoka’s Labs at UW - Madison.
Motivation: Virus-Host Interaction

Which genes does the virus hijack for reproduction?
Basic Model

Genome

Hijacked Proteins

Discovered Hijacked Genes
Basic Model

Genome

Knockdown

Hijacked Proteins

Discovered Hijacked Genes
Basic Model

Genome

Hijacked Proteins

Discovered Hijacked Genes
Basic Model

Genome

Knockdown

Hijacked Proteins

Discovered Hijacked Genes
Basic Model

Genome

Knockdown

Hijacked Proteins

Discovered Hijacked Genes
Basic Model

Genome

Hijacked Proteins

Discovered Hijacked Genes

13,071 genes to check in a fruit fly.
Redundant Model

Genome

Hijacked Proteins

Discovered Hijacked Genes
Redundant Model

Genome

Knockdown

Hijacked Proteins

Discovered Hijacked Genes
Redundant Model

Genome

Knockdown

Knockdown

Hijacked Proteins

Discovered Hijacked Genes
Redundant Model

Genome

Knockdown  Knockdown

Hijacked Proteins

Discovered Hijacked Genes
Redundant Model

Genome:

Knockdown:

Hijacked Proteins:

Discovered Hijacked Genes:
Redundant Model

Genome

Hijacked Proteins

Discovered Hijacked Genes
Redundant Model

Genome

Knockdown
Knockdown

Hijacked Proteins

Discovered Hijacked Genes

Millions of pairs of genes to check in a fruit fly.
Can model output (virus expression) as a sparse linear system.

\[ y = \sum_{i} a_i x_i \]

\( x_i \) corresponds to whether gene is involved or not.

\( a_i \) corresponds to knockdowns.
Sparse Multilinear Systems

Genome

Knockdown

Knockdown

\[ y = \sum_{i<j} a_i a_j x_{ij} \]

Can model output (virus expression) as a sparse multilinear system.

\( x_i \) corresponds to whether gene is involved or not.

\( a_i \) corresponds to knockdowns.
Sparse Multilinear Systems

Can model output (virus expression) as a sparse multilinear system.

\[ y = \sum_{i} a_i x_i + \sum_{i<j} a_i a_j x_{ij} \]

- \( x_i \) corresponds to whether gene is involved or not.
- \( a_i \) corresponds to knockdowns.
This Talk

How many measurements are needed to infer the coefficients for a sparse multilinear system?

\[ y = \sum_{i_1} a_{i_1} x_{i_1} + \sum_{i_1 < i_2} a_{i_1} a_{i_2} x_{i_1 i_2} + \cdots + \sum_{i_1 < i_2 < \cdots < i_D} a_{i_1} a_{i_2} \cdots a_{i_D} x_{i_1 i_2 \cdots i_D} \]
This Talk

How many measurements are needed to infer the coefficients for a sparse multilinear system?

\[ y = \sum_{i_1} a_{i_1} x_{i_1} + \sum_{i_1 < i_2} a_{i_1} a_{i_2} x_{i_1 i_2} + \cdots \]

\[ \cdots + \sum_{i_1 < i_2 < \cdots < i_D} a_{i_1} a_{i_2} \cdots a_{i_D} x_{i_1 i_2 \cdots i_D} \]

\[ \downarrow \]

Sufficient to consider order D interactions

\[ y = \sum_{i_1 < i_2 < \cdots < i_D} a_{i_1} a_{i_2} \cdots a_{i_D} x_{i_1 i_2 \cdots i_D} \]
Vectorized Multilinear System

\[ y_k = \sum_{i_1 < i_2 < \cdots < i_D} a_{ki_1} a_{ki_2} \cdots a_{ki_D} x_{i_1 i_2 \cdots i_D} \]

( Take inputs to be binary symmetric for simplicity. )
Vectorized Multilinear System

\[ y_k = \sum_{i_1 < i_2 < \cdots < i_D} a_{k i_1} a_{k i_2} \cdots a_{k i_D} x_{i_1 i_2 \cdots i_D} \]

( Take inputs to be binary symmetric for simplicity. )

Write coefficients as a vector
Vectorized Multilinear System

\[ y_k = \sum_{i_1<i_2<\cdots<i_D} a_{ki_1} a_{ki_2} \cdots a_{ki_D} x_{i_1 i_2 \cdots i_D} \]

(Take inputs to be binary symmetric for simplicity.)

Write coefficients as a vector:
\[ \{a_{ki_1} a_{ki_2} \cdots a_{ki_D}\} \Rightarrow \mathbf{a}_k \]

Write measurements as a vector:
\[ \{x_{i_1 i_2 \cdots i_D}\} \Rightarrow \mathbf{x} \]
Vectorized Multilinear System

\[ y_k = \sum_{i_1 < i_2 < \cdots < i_D} a_{k_1} a_{k_2} \cdots a_{k_D} x_{i_1 i_2 \cdots i_D} \]

(Take inputs to be binary symmetric for simplicity.)

\[ \{ x_{i_1 i_2 \cdots i_D} \} \xrightarrow{\text{Write coefficients as a vector}} x \]

\[ \{ a_{k_1} a_{k_2} \cdots a_{k_D} \} \xrightarrow{\text{Write measurements as a vector}} a_k \]

Back to a linear problem: \( y = Ax \) with potentially dependent measurements.
Vectorized Multilinear System

\[ y_k = \sum_{i_1 < i_2 < \cdots < i_D} a_{ki_1} a_{ki_2} \cdots a_{ki_D} x_{i_1 i_2 \cdots i_D} \]

(Take inputs to be binary symmetric for simplicity.)

\[
\begin{align*}
\{x_{i_1 i_2 \cdots i_D}\} & \quad \text{Write coefficients as a vector} \\
\{a_{ki_1} a_{ki_2} \cdots a_{ki_D}\} & \quad \text{Write measurements as a vector}
\end{align*}
\]

Back to a linear problem: \( y = Ax \) with potentially dependent measurements.

Can we get down to \( S \log \left( \frac{N}{S} \right) \) measurements?
Compressed Sensing

Candes-Romberg-Tao ’06, Candes-Tao ’06, Donoho ’06

Linear measurements:

\[
y_k = \sum_{i} a_{ki} x_i = a_k^T x
\]

Coefficients are S-sparse:

\[
\|x\|_0 = S
\]

\[
x \in \mathbb{R}^N
\]
Compressed Sensing

Candes-Romberg-Tao ’06, Candes-Tao ’06, Donoho ’06

Linear measurements:

\[ y_k = \sum_i a_{ki} x_i = a_k^T x \]
\[ y = Ax \]

Coefficients are S-sparse:

\[ \| x \|_0 = S \]
\[ x \in \mathbb{R}^N \]

If the measurement matrix satisfies the restricted isometry property (RIP) with \( \delta_{2S} < \sqrt{2} - 1 \) for all S-sparse vectors:

\[ (1 - \delta_S) \| x \|_2^2 \leq \| Ax \|_2^2 \leq (1 + \delta_S) \| x \|_2^2 \]
Compressed Sensing
Candes-Romberg-Tao ’06, Candes-Tao ’06, Donoho ’06

Linear measurements:

\[ y_k = \sum_i a_{ki} x_i = a_k^T x \]
\[ y = Ax \]

Coefficients are S-sparse:

\[ \|x\|_0 = S \]
\[ x \in \mathbb{R}^N \]

If the measurement matrix satisfies the restricted isometry property (RIP) with \( \delta_{2S} < \sqrt{2} - 1 \) for all S-sparse vectors:

\[ (1 - \delta_S)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_S)\|x\|_2^2 \]

Then we can recover the coefficients through an \( \ell_1 \) optimization:

\[ \min \|\hat{x}\|_1 \text{ subject to } A\hat{x} = y \]
Compressed Sensing

Candes-Romberg-Tao ’06, Candes-Tao ’06, Donoho ’06

Linear measurements:

\[ y_k = \sum_i a_{ki} x_i = a_k^T x \]
\[ y = Ax \]

Coefficients are S-sparse:

\[ \| x \|_0 = S \]
\[ x \in \mathbb{R}^N \]

If the measurement matrix satisfies the restricted isometry property (RIP) with \( \delta_{2S} < \sqrt{2} - 1 \) for all S-sparse vectors:

\[
(1 - \delta_S) \| x \|_2^2 \leq \| A x \|_2^2 \leq (1 + \delta_S) \| x \|_2^2
\]

Then we can recover the coefficients through an \( \ell_1 \) optimization:

\[
\min \| \hat{x} \|_1 \text{ subject to } A \hat{x} = y
\]

Can get RIP with \( K \geq cS \log(N/S) \) measurements.
Norm Preservation

The restricted isometry property guarantees that the norms of all sparse vectors are approximately preserved.

Example: $M = 3$ inputs, $D = 2$ interactions \[ \mathbf{a} = [a_1 a_2 \ a_1 a_3 \ a_2 a_3] \]
Norm Preservation

The restricted isometry property guarantees that the norms of all sparse vectors are approximately preserved.

Example: \( M = 3 \) inputs, \( D = 2 \) interactions \( \mathbf{a} = [a_1 a_2 \ a_1 a_3 \ a_2 a_3] \)

Our measurements preserve the norm in expectation:

\[
\mathbb{E} \left[ \| \mathbf{y} \|^2 \right] = \mathbf{x}^T \mathbb{E} \left[ \mathbf{A}^T \mathbf{A} \right] \mathbf{x}
\]

Each diagonal element is the sum of products of squared terms.
Each off-diagonal element is the sum of products, with at least one unique term.

\[
\mathbb{E} \left[ \mathbf{A}^T \mathbf{A} \right] = \mathbf{I}
\]

\[
\mathbb{E} \left[ \| \mathbf{y} \|^2 \right] = \| \mathbf{x} \|^2
\]
Example: Best Case Sparsity Pattern

If each index is involved in at most one interaction, then the measurements concentrate very quickly.

Example: \[ y = a_1 a_2 x_{12} + a_3 a_5 x_{35} + a_4 a_7 x_{47} \]
If each index is involved in at most one interaction, then the measurements concentrate very quickly.

Example: \[ y = a_1 a_2 x_{12} + a_3 a_5 x_{35} + a_4 a_7 x_{47} \]

Relabel \[ y = \tilde{a}_1 \tilde{x}_1 + \tilde{a}_2 \tilde{x}_2 + \tilde{a}_3 \tilde{x}_3 \]
Example: Best Case Sparsity Pattern

If each index is involved in at most one interaction, then the measurements concentrate very quickly.

Example: \[ y = a_1 a_2 x_{12} + a_3 a_5 x_{35} + a_4 a_7 x_{47} \]

Relabel

\[ y = \tilde{a}_1 \tilde{x}_1 + \tilde{a}_2 \tilde{x}_2 + \tilde{a}_3 \tilde{x}_3 \]

Best case concentration is subgaussian:

\[ \inf_{T} \mathbb{P}(|y_k^2 - 1| > t) \leq \exp(-ct) \]
If each a subset of indices is involved in event interaction, then the measurements concentrate more slowly.

Example: \( y = \sum_{1 \leq i \leq j \leq \sqrt{S}} a_i a_j \) \[ \mathbb{P} \left( y^2 \geq S \right) \geq 2^{-\sqrt{S}} \]

Worst case concentration is subexponential:

\[
\sup_{\mathcal{T}} \mathbb{P}(\left| y_k^2 - 1 \right| > t) \geq \exp(-ct^{1/D})
\]
Eigenvalue Experiment

Look at minimum eigenvalues of subsets of Gram matrix for order $D = 1, 2, 3$ interactions.

Obviously, there is some cost incurred by dependencies within the vector.
Gershgorin’s Disc Theorem

Approach inspired by Haupt-Bajwa-Raz-Nowak ’08:

i) Control each element of the Gram matrix \( G_R = A_R^T A_R \) via Hoeffding’s inequality. Bound probability that Gram matrix is at worst:

\[
G_R = \begin{bmatrix}
1 & \frac{\delta_S}{S} & \cdots & \frac{\delta_S}{S} \\
\frac{\delta_S}{S} & 1 & \cdots & \frac{\delta_S}{S} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\delta_S}{S} & \frac{\delta_S}{S} & \cdots & 1
\end{bmatrix}
\]
Gershgorin's Disc Theorem

Approach inspired by Haupt-Bajwa-Raz-Nowak ’08:

i) Control each element of the Gram matrix \( \mathbf{G}_\mathcal{R} = \mathbf{A}_\mathcal{R}^T \mathbf{A}_\mathcal{R} \) via Hoeffding's inequality. Bound probability that Gram matrix is at worst:

\[
\mathbf{G}_\mathcal{R} = \begin{bmatrix}
1 & \frac{\delta_S}{S} & \cdots & \frac{\delta_S}{S} \\
\frac{\delta_S}{S} & 1 & \cdots & \frac{\delta_S}{S} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\delta_S}{S} & \frac{\delta_S}{S} & \cdots & 1
\end{bmatrix}
\]

ii) Gershgorin’s Disc Theorem guarantees that eigenvalues lie in the range:

\[
g_{ii} - \sum_{j \neq i} |g_{ij}| \leq \lambda_i(\mathbf{G}_\mathcal{R}) \leq g_{ii} + \sum_{j \neq i} |g_{ij}|
\]
Gershgorin’s Disc Theorem

Approach inspired by Haupt-Bajwa-Raz-Nowak ’08:

i) Control each element of the Gram matrix $G_R = A_R^T A_R$ via Hoeffding’s inequality. Bound probability that Gram matrix is at worst:

$$G_R = \begin{bmatrix}
1 & \frac{\delta_S}{S} & \ldots & \frac{\delta_S}{S} \\
\frac{\delta_S}{S} & 1 & \ldots & \frac{\delta_S}{S} \\
\vdots & \vdots & \ddots & \vdots \\
\frac{\delta_S}{S} & \frac{\delta_S}{S} & \ldots & 1
\end{bmatrix}$$

ii) Gershgorin’s Disc Theorem guarantees that eigenvalues lie in the range:

$$g_{ii} - \sum_{j \neq i} |g_{ij}| \leq \lambda_i(G_R) \leq g_{ii} + \sum_{j \neq i} |g_{ij}|$$

iii) Union bound over all $\binom{N}{S}$ sparse patterns. Get RIP with constant $\delta_S$ long as number of measurements is at least:

$$K = c \ S^2 \log N$$
Heavy-Tailed Restricted Isometries

Rudelson-Vershynin ’08, Vershynin ’10:
Assume the expected Gram matrix is identity, \( \mathbb{E} [A^T A] = I \), and the rows of the measurement matrix are independent, then if the number of measurements is at least

\[
K = c \frac{S}{\epsilon^2} \left( \log \left( \frac{S}{\epsilon^2} \right) \right)^3 \log N
\]

with expected RIP constant

\( \mathbb{E} [\delta_S] \leq \epsilon \)
Heavy-Tailed Restricted Isometries

Rudelson-Vershynin ’08, Vershynin ’10:
Assume the expected Gram matrix is identity, $\mathbb{E} [A^T A] = I$, and the rows of the measurement matrix are independent, then if the number of measurements is at least

$$K = c \frac{S}{\epsilon^2} \left( \log \left( \frac{S}{\epsilon^2} \right) \right)^3 \log N$$

with expected RIP constant $\mathbb{E} [\delta_S] \leq \epsilon$

One of the key points in the proof is that it avoids using a union bound and instead bounds the RIP constants of all sparsity patterns simultaneously.
Heavy-Tailed Restricted Isometries

Rudelson-Vershynin ’08, Vershynin ’10:
Assume the expected Gram matrix is identity, $\mathbb{E} \left[ A^T A \right] = I$, and the rows of the measurement matrix are independent, then if the number of measurements is at least

$$K = c \frac{S}{\epsilon^2} \left( \log \left( \frac{S}{\epsilon^2} \right) \right)^3 \log N$$

with expected RIP constant

$$\mathbb{E} [\delta_S] \leq \epsilon$$

One of the key points in the proof is that it avoids using a union bound and instead bounds the RIP constants of all sparsity patterns simultaneously.

This corresponds exactly to our linearized problem.
How exactly do the tails behave?

Rademacher Chaos of order D:

\[ y = \sum_{i_1 < i_2 < \cdots < i_D} a_{i_1} a_{i_2} \cdots a_{i_D} x_{i_1 i_2 \cdots i_D} \]

\(a_i\) are independent and binary symmetric (Rademacher).
How exactly do the tails behave?

Rademacher Chaos of order \( D \):

\[
y = \sum_{i_1 < i_2 < \cdots < i_D} a_{i_1} a_{i_2} \cdots a_{i_D} x_{i_1 i_2 \cdots i_D}
\]

\( a_i \) are independent and binary symmetric (Rademacher).

The **combinatorial dimension** measures the level of dependence introduced by the sparsity pattern \( \mathcal{T}_L \). A chaos has combinatorial dimension \( \alpha \) if

\[
\sup_{A_1, A_2, \ldots, A_D} \frac{|\mathcal{T}_L \cap (A_1 \times A_2 \times \cdots \times A_D)|}{(\max_{1 \leq j \leq D} |A_j|)^{\alpha}} \leq C_1
\]

\[
|\mathcal{T}_L| \geq C_2 L^\alpha
\]

Takes values between \( 1 \leq \alpha \leq D \)
Chaos Tail Bounds

Blei-Janson ’04: A Rademacher chaos with combinatorial dimension $\alpha$ satisfies:

$$\exp \left( -c_1 \frac{t^2}{\alpha} \right) \leq \sup_{L} \mathbb{P} (|u| > t) \leq \exp \left( -c_2 \frac{t^2}{\alpha} \right)$$
Chaos Tail Bounds

Blei-Janson ’04: A Rademacher chaos with combinatorial dimension $\alpha$ satisfies:

$$\exp \left( -c_1 t^{2/\alpha} \right) \leq \sup_L \mathbb{P} \left( |u| > t \right) \leq \exp \left( -c_2 t^{2/\alpha} \right)$$

Using this bound it can be shown that:

$$\mathbb{P} \left( \left| \|y\|^2 - 1 \right| \right) \leq \exp \left( -c \max \left( K t / S, K^{1/\alpha} t^{1/\alpha} \right) \right)$$

from which we can derive the number of measurements needed to get RIP for a single pattern.
Chaos Tail Bounds

Blei-Janson '04: A Rademacher chaos with combinatorial dimension $\alpha$ satisfies:

$$\exp \left( -c_1 t^{2/\alpha} \right) \leq \sup_L \mathbb{P} (|u| > t) \leq \exp \left( -c_2 t^{2/\alpha} \right)$$

Using this bound it can be shown that:

$$\mathbb{P} \left( \left\| y \right\|^2 - 1 \right) \leq \exp \left( -c \max \left( Kt/S, K^{1/\alpha} t^{1/\alpha} \right) \right)$$

from which we can derive the number of measurements needed to get RIP for a single pattern.

Union bound technique from Baraniuk-Devore-Davenport-Wakin '08:

$$K \geq \min \left( c \ S^2 \log(N/S), \ c \ S^\alpha \log^\alpha(N/S) \right)$$
<table>
<thead>
<tr>
<th>Method</th>
<th>Bound</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gershgorin</td>
<td>$S^2 \log N$</td>
<td>Second Moment Union Bound</td>
</tr>
<tr>
<td>Rudelson-Vershynin</td>
<td>$S(\log^3 S)(\log N)$</td>
<td>Second Moment No Union Bound</td>
</tr>
<tr>
<td>Rademacher Chaos</td>
<td>$S^\alpha \log^\alpha (N/S)$</td>
<td>Tail Bounds Union Bound</td>
</tr>
</tbody>
</table>
Conclusions and Future Directions

Can recover sparse vectors from multilinear systems despite heavy-tailed behavior and dependencies.

Number of measurements may depend on the pattern of sparsity.

What is possible for polynomial systems?
Acknowledgments

Thanks to Paul Ahlquist for introducing us to virus replication.

Thanks to Ben Recht for introducing us to the work of Rudelson and Vershynin.