(Semi-)Algebraic Geometry at IBM Research – Ireland

Jakub Mareček
IBM Research – Ireland

with Tianran Chen, Bissan Ghaddar, Allan C. Liddell, Jie Liu, Timothy McCoy, Dhagash Mehta, Martin Mevissen, Matthew Niemerg, and Martin Takáč
Acknowledgements

With much gratitude to my co-authors:

- Tianran Chen (Auburn University)
- Bissan Ghaddar (University of Waterloo)
- Jie Liu and Martin Takáč (Lehigh University)
- Allan C. Liddell, Timothy McCoy, and Dhagash Mehta (Notre Dame)
- Martin Mevissen and Matthew Niemerg (IBM Research)
- Jonas C. Villumsen (Hitashi)
- Guanglei Wang (Pierre and Marie Curie University)

Cf.:

- JM, McCoy, Mevissen, arXiv:1412.8054
- Chen, JM, Mehta, Niemerg, to appear
- Ghaddar, JM, Mevissen, arXiv:1404.3626
- JM, Takáč: arXiv:1506.08568
- Liu, JM, Takáč: arXiv:1510.02171
The Agenda

An Introduction
Some structural results
Convexifications based on semidefinite programming (SDP)
SDP solvers based on tree-width decompositions, $r$-space sparsity, and interior-point methods
SDP solvers based on coordinate descent
Solvers combining a convexification with Newton method on the non-convex problem
IBM Research

- One of the world’s largest corporate research labs
- Driving 23 consecutive years of IBM patent leadership
- 6 Nobel laureates, 6 Turing awards, etc.
- A diversity of core academic disciplines, incl:
  - Computer science
  - Mathematical sciences
  - Behavioural science
  - Chemistry, Physics, Materials science
IBM Research – Ireland

- 1 of 12 IBM Research labs globally
- 100+ employees, plus a UCD co-lab
- 30 patents and over 100 top-tier publications in 2015
- IoT: Monitoring, forecasting, optimization, control

Highlights:
- A large body of work on power systems
- Work with water authorities across 3 continents
- Work with transportation authorities across 3 continents
- Work with airlines and a major vendor of software for airlines
- Work with a number of major car manufacturers
The Mathematical Challenges

- Power systems exhibit non-linear dynamics, which are uncertain (human behaviour) and time-varying (demand)
- Water systems exhibit non-linear dynamics, which are uncertain (human behaviour) and time-varying (demand)
- Transportation exhibits non-linear dynamics, which are uncertain (human behaviour) and time-varying (demand)
- In general, control of non-linear systems in undecidable.
The Mathematical Challenges

- An abstraction of the dynamics in power systems is the differential equation:

\[
\frac{d\theta_i}{dt} = \omega_i - \frac{1}{N}\sum_{j=1}^{N} K_{i,j} \sin(\theta_i - \theta_j), \text{ for } i = 1, \ldots, N, \tag{1}
\]

where \(N\) is the number of oscillators, \(K_{i,j}\) is the coupling strength between the \(i\)-th and \(j\)-th oscillators. The matrix \(K = [K_{i,j}]\) may also be viewed as the adjacency matrix for the underlying weighted graph. \(\Omega = (\omega_1, \ldots, \omega_N)\) contains the natural frequencies of the \(N\) oscillators.
The Mathematical Challenges

- Within the phase space of this system of ODEs are the steady states, where $\frac{d\theta_i}{dt} = 0$ for all $i = 1, \ldots, N$.
- An analysis of the equilibria can reveal the behavior of the dynamical system near the equilibria.
- Even for the steady-state problems, which are solved every couple of minutes, there are no good solvers.
- How bad can this be?
## A FERC Study: Where do Leading Solvers Fail?

<table>
<thead>
<tr>
<th>Termination</th>
<th>Conopt</th>
<th>Ipopt</th>
<th>Knitro</th>
<th>Minos</th>
<th>Snopt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exceeded time limit with an infeasible solution</td>
<td>7.9%</td>
<td>22.2%</td>
<td>38.4%</td>
<td>0.0%</td>
<td>23.0%</td>
</tr>
<tr>
<td>Exceeded time limit with a feasible solution</td>
<td>6.5%</td>
<td>0.9%</td>
<td>6.7%</td>
<td>0.0%</td>
<td>14.4%</td>
</tr>
<tr>
<td>Early termination with an infeasible solution</td>
<td>0.0%</td>
<td>0.4%</td>
<td>0.6%</td>
<td>11.4%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Early termination with a feasible solution</td>
<td>31.2%</td>
<td>0.6%</td>
<td>1.8%</td>
<td>2.3%</td>
<td>0.2%</td>
</tr>
<tr>
<td>Normal termination with an infeasible solution</td>
<td>10.5%</td>
<td>3.4%</td>
<td>0.0%</td>
<td>55.9%</td>
<td>1.6%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Percentage of feasible</th>
<th>Percentage of best seen</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>81.6%</td>
<td>74.1%</td>
</tr>
<tr>
<td></td>
<td>61.0%</td>
<td>52.5%</td>
</tr>
<tr>
<td></td>
<td>32.7%</td>
<td>30.3%</td>
</tr>
<tr>
<td></td>
<td>75.4%</td>
<td>60.8%</td>
</tr>
</tbody>
</table>

Source: Anya Castillo and Richard O’Neill, Staff Technical Conference, FERC, Washington DC, June 24–26, 2013. 118–3120 bus instances, 100 runs per instance and solver, 20 min runs, Xeon E7458 2.4GHz, 64 GB RAM
The Agenda

An Introduction

**Some structural results**
Convexifications based on semidefinite programming (SDP)
SDP solvers based on tree-width decompositions, $r$-space sparsity, and interior-point methods
SDP solvers based on coordinate descent
Solvers combining a convexification with Newton method on the non-convex problem
Motivational Eye Candy

Figure: Smooth convex

Figure: Smooth nonconvex

Figure: Nonsmooth convex

Figure: Nonsmooth nonconvex
The Steady State Problems

Real-valued:

- Polar Power-Voltage \( p_k, q_k, v_k, \theta_k \)
- Rectangular Power-Voltage (R-PQV) \( p_k, q_k, \Re v_k, \Im v_k \)
- Polar Current Injection \( p_k, q_k, v_k, \theta_k, i_k \)
- Rectangular Current Injection (R-CI) \( p_k, q_k, \Re v_k, \Im v_k, i_k \)
- Polar Voltage-Current \( p_k, q_k, v_k, \theta_k, i_{lm} \)
- Rectangular Voltage-Current (R-IV) \( p_k, q_k, \Re v_k, \Im v_k, i_{lm} \)

Further variants: dropping \( p_k, q_k \), mixing polar and rectangular.

Complex-valued:

- With complex conjugate (PQV, CI, IV)
- Without (McCoy’s trick)

Rectangular Power-Voltage Formulation (PQV)

Variables:

- $v_n$ is complex voltage at bus $n$
- $p_n$ is active power at bus $n$ (the real part)
- $q_n$ is reactive power at bus $n$ (imag. part)

Key constraints:

\[
\Re v_n \sum_{m \in N} (g_{nm}\Re v_m - b_{nm}\Im v_m) + \Im v_n \sum_{m \in N} (g_{nm}\Im v_m + b_{nm}\Re v_m) - p_n + p_n^d = 0
\]
\[
\Im v_n \sum_{m \in N} (g_{nm}\Re v_m - b_{nm}\Im v_m) + \Re v_n \sum_{m \in N} (g_{nm}\Im v_m + b_{nm}\Re v_m) - q_n + q_n^d = 0
\]
PQV: Derivation from the Dynamics

- Using the identities \(\sin(\theta_i - \theta_j) = \sin \theta_i \cos \theta_j - \sin \theta_j \cos \theta_i\)
- Consider the substitution \(s_i := \sin \theta_i\) and \(c_i := \cos \theta_i\) for all \(i = 1, \ldots, N - 1\), and \(s_i^2 + c_i^2 - 1 = 0\), for all \(i = 1, \ldots, N - 1\)
- A system of polynomial equations:

\[
\omega_i - \frac{1}{N} \sum_{j=1}^{N} K_{i,j} (s_i c_j - s_j c_i) = 0
\]

\[
s_i^2 + c_i^2 - 1 = 0,
\]

for \(i = 1, \ldots, N - 1\).
The Structural Results

- A reformulation of the complex conjugate
- A \( \binom{2(N-1)}{N-1} \) bound on the number of feasible solutions considering losses, improving upon Bézout bound of \( 2^{2(N-1)} \), and resolving a conjecture of Baillieul and Byrnes, which has been open for over three decades
- A variety of additional structural results based on a reformulation of the steady-state equations to a multi-homogeneous algebraic system
- A bound on the number of feasible solutions considering losses and the structure of the power system, based on the work of Bernstein
- Illustrations on some well-known instances, including the numbers of roots, conditions for non-uniqueness of optima, and tree-width.
The Multi-Homogeneous Structure

- Replace all $v_n^*$ with independent variables $u_n$,
- Filter for solutions where $u_n = v_n^*$ once the complex solutions are obtained.
- Variables $v_n$ and $u_n$ produce a multi-homogeneous structure with variable groups $\{v_n\}$ and $\{u_n\}$:

$$
\begin{align*}
    v_n \sum_k Y_{n,k} u_k + u_n \sum_k Y_{n,k}^* v_k &= 2p_n & n \in N \setminus G \\
    v_n \sum_k Y_{n,k} u_k - u_n \sum_k Y_{n,k}^* v_k &= 2q_n & n \in N \setminus G \\
    v_n u_n &= |v_n|^2 & n \in G - \{0\} \\
    v_0 &= |v_0|, \quad u_0 = |v_0|
\end{align*}
$$

where $G$ be the set of slack generators for which $|v_n|$ is specified, and $0 \in G$ corresponds to a reference node with phase $0$. 

(3)
For example for the two-bus network, we obtain:

\[\nu_1(Y_{1,0}u_0 + Y_{1,1}u_1) + u_1(Y_{1,0}^*\nu_0 + Y_{1,1}^*\nu_1) = 2p_1\]
\[\nu_1(Y_{1,0}u_0 + Y_{1,1}u_1) - u_1(Y_{1,0}^*\nu_0 + Y_{1,1}^*\nu_1) = 2q_1\]
\[\nu_0 = |\nu_0|, \ u_0 = |\nu_0| \] (4)
The Structural Results

**Theorem**

With exceptions on a parameter set of measure zero, the alternating-current power flow (1) has a finite number of complex solutions, which is bounded above by:

$$\binom{2n - 2}{n - 1}$$

(5)
The Structural Results

Proof.

Each equation in the system (1) is linear in the $V$ variables and also in the $V^*$ variables, giving rise to a natural multi-homogeneous structure of multi-idegree $(1, 1)$. Since the slack bus voltage is fixed at a reference value, the system has $2n - 2$ such equations in $(n - 1, n - 1)$ variables. By the multi-homogeneous form of Bézout’s Theorem, the total number of solutions in multi-projective space $\mathbb{CP}^{n-1} \times \mathbb{CP}^{n-1}$ is precisely the stated bound, counting multiplicity.
## An Illustration

The maximum number of steady states in a circuit with a fixed number of buses:

<table>
<thead>
<tr>
<th>Method</th>
<th>No.</th>
<th>N of buses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Bézout’s upper bound</td>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td>A BKK-based upper bound</td>
<td>8</td>
<td>40</td>
</tr>
<tr>
<td>Theorem 1</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>Generic lower bound</td>
<td>6</td>
<td>20</td>
</tr>
</tbody>
</table>
The Structural Results

Theorem (Theorem 1 of Malajovich and Meer)

*There does not exist a polynomial time algorithm to approximate the minimal multi-homogeneous Bézout number for a polynomial system up to any fixed factor, unless \( P = NP \).*
The Structural Results

Corollary

If there exists a feasible solution of the alternating-current power flow, then the solution has even multiplicity greater or equal to 2 or another solution exists.
The Theorem (Bernstein 1975)

Given a system of $n$ polynomial equations in $n$ variables, the number of isolated complex solutions for which no variable is zero is bounded above by the mixed volume of the Newton polytopes of the equations.
Alternative Derivations from the Dynamics

• Using the trigonometric identity
  \[ \sin(\theta_i - \theta_j) = \frac{1}{2\ii}(e^{\ii(\theta_i - \theta_j)} - e^{-\ii(\theta_i - \theta_j)}) \]
  where \( \ii := \sqrt{-1} \) is the imaginary unit.

• Introduce substitution:
  \[ x_i := e^{\ii \theta_i} \quad \text{and} \quad y_i := e^{-\ii \theta_i}, \]
  for all \( i = 1, \ldots, N - 1 \).

• An enlarged system of \( 2(N - 1) \) equations in \( 2(N - 1) \) variables
  \[ \sum_{j=1}^{N} \frac{K_{i,j}}{\II N} (x_i y_j - x_j y_i) = \omega_i \quad \text{for} \quad i = 1, \ldots, N - 1 \]
  \( x_i y_i = 1 \quad \text{for} \quad i = 1, \ldots, N - 1. \) (6)
The Structural Results

For example, with $N = 3$, the system becomes

\[
\begin{align*}
\frac{K_{1,2}}{3\pi}(x_1 y_2 - x_2 y_1) + \frac{K_{1,3}}{3\pi}(x_1 - y_1) - \omega_1 &= 0 \\
\frac{K_{2,1}}{3\pi}(x_2 y_1 - x_1 y_2) + \frac{K_{2,3}}{3\pi}(x_2 - y_2) - \omega_2 &= 0
\end{align*}
\]

(7)

\[x_1 y_1 - 1 = 0\]

\[x_2 y_2 - 1 = 0.\]
The Structural Results

• Yet another transformation based on identities

\[
\sin \theta_i = \frac{2 \tan \frac{\theta_i}{2}}{1 + \tan^2 \frac{\theta_i}{2}} \quad \text{and} \quad \cos \theta_i = \frac{1 - \tan \frac{\theta_i}{2}}{1 + \tan^2 \frac{\theta_i}{2}},
\]

(8)

• One introduces \( t_i := \tan \frac{\theta_i}{2} \).
The Structural Results

For example, with $N = 3$, the system becomes

$$K_{1,2} \left( \frac{2t_1}{1 + t_1^2} \frac{1 - t_2}{1 + t_2^2} - \frac{2t_2}{1 + t_2^2} \frac{1 - t_1}{1 + t_1^2} \right) + K_{1,3} \left( \frac{2t_1}{1 + t_1^2} \frac{1 - t_3}{1 + t_3^2} - \frac{2t_3}{1 + t_3^2} \frac{1 - t_1}{1 + t_1^2} \right) = 3\omega_1$$

$$K_{2,1} \left( \frac{2t_2}{1 + t_2^2} \frac{1 - t_1}{1 + t_1^2} - \frac{2t_1}{1 + t_1^2} \frac{1 - t_2}{1 + t_2^2} \right) + K_{2,3} \left( \frac{2t_2}{1 + t_2^2} \frac{1 - t_3}{1 + t_3^2} - \frac{2t_3}{1 + t_3^2} \frac{1 - t_2}{1 + t_2^2} \right) = 3\omega_1$$

considering $t_i := \tan \frac{\theta_i}{2}$. Subsequently, one increases the degree by multiplication.
The Structural Results

**Theorem**

If there exists a choice of $K_{i,j}$’s and $\omega_i$’s for which the number of nonzero complex solutions of (6) is the BKK bound, then for almost all choices of complex $K_{i,j}$’s and $\omega_i$, the number of nonzero complex solutions of (6) will be the BKK bound.
## An Illustration on the Path Graph

<table>
<thead>
<tr>
<th>Nodes</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bezout</td>
<td>16</td>
<td>64</td>
<td>256</td>
<td>1024</td>
<td>4096</td>
<td>16384</td>
<td>65536</td>
<td>262144</td>
</tr>
<tr>
<td>Binomial</td>
<td>6</td>
<td>20</td>
<td>70</td>
<td>252</td>
<td>924</td>
<td>3432</td>
<td>12870</td>
<td>48620</td>
</tr>
<tr>
<td>BKK for (2)</td>
<td>8</td>
<td>24</td>
<td>80</td>
<td>256</td>
<td>832</td>
<td>2688</td>
<td>8704</td>
<td>28160</td>
</tr>
<tr>
<td>BKK for (6)</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>32</td>
<td>64</td>
<td>128</td>
<td>256</td>
<td>512</td>
</tr>
<tr>
<td>Generic root count</td>
<td>4</td>
<td>8</td>
<td>16</td>
<td>32</td>
<td>64</td>
<td>128</td>
<td>256</td>
<td>512</td>
</tr>
</tbody>
</table>
More Structural Results

**Remark**

For the alternating-current power flows, where powers are fixed at all but the reference bus, whenever there exists a real feasible solution, except for a parameter set of measure zero, one can enumerate all feasible solutions in finite time.
Computations with Bertini

variable_group V0, V1; variable_group U0, U1;
I0 = V0*Yv0_0 + V1*Yv0_1; I1 = V0*Yv1_0 + V1*Yv1_1;
J0 = U0*Yu0_0 + U1*Yu0_1; J1 = U0*Yu1_0 + U1*Yu1_1;
fv0 = V0 - 1.0; fv1 = I1*U1 + J1*V1 + 7.0;
fu0 = U0 - 1.0; fu1 = -I1*U1 + J1*V1 - 7.0*I;

A Bertini encoding of ACPF on the two-bus instance of Bukhsh et al. 2012, where the impedance of a single branch is 0.04 + 0.2i.
Properties of the Steady States

- One can optimise a variety of objectives over the steady states. The costs of real power $P_0$ generated at the reference bus 0 by a quadratic function $f_0$:

$$\text{cost} := f_0(P_0).$$  \hspace{1cm} (9)

- A norm of the vector $D$ obtained by summing apparent powers $S(u, v) + S(v, u) \forall (u, v) \in E$ for:

$$\|D\|_p = \left( \sum_{(u, v) \in E} |S(u, v) + S(v, u)|^p \right)^{1/p}. \hspace{1cm} (10)$$

- The usual $\|D\|_1$ is denoted loss below.
An Illustration

<table>
<thead>
<tr>
<th>Instance</th>
<th>Source</th>
<th>No. of buses</th>
<th>No. of branches</th>
<th>Treewidth tw(P)</th>
<th>No. of solutions</th>
<th>$\min_{x \in X} (\text{cost}(x))$</th>
<th>No. of min. wrt. cost</th>
<th>$\max_{x \in X} (\text{cost}(x))$</th>
<th>$\min_{x \in X} (\text{loss}(x))$</th>
<th>No. of min. wrt. loss</th>
<th>$\max_{x \in X} (\text{loss}(x))$</th>
</tr>
</thead>
<tbody>
<tr>
<td>case2w</td>
<td>Bukhsh et al. [?]</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>8.42</td>
<td>1</td>
<td>9.04</td>
<td>9.66</td>
<td>0.71</td>
<td></td>
</tr>
<tr>
<td>case3KW</td>
<td>Klos and Wojcicka [?]</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>6</td>
<td>-0.0</td>
<td>1</td>
<td>1250.0</td>
<td>1500.0</td>
<td>-0.0</td>
<td></td>
</tr>
<tr>
<td>case3LL</td>
<td>Lavaei and Low [?]</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1502.07</td>
<td>1</td>
<td>1502.19</td>
<td>1502.31</td>
<td>0.22</td>
<td></td>
</tr>
<tr>
<td>case3w</td>
<td>Bukhsh et al. [?]</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5.88</td>
<td>1</td>
<td>5.94</td>
<td>6.01</td>
<td>0.55</td>
<td></td>
</tr>
<tr>
<td>case4</td>
<td>McCoy et al.</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>1502.51</td>
<td>1</td>
<td>1505.19</td>
<td>1507.88</td>
<td>0.11</td>
<td></td>
</tr>
<tr>
<td>case4ac</td>
<td>McCoy et al.</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>2005.25</td>
<td>2</td>
<td>3337.5</td>
<td>4005.51</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>case4cyc</td>
<td>Bukhsh et al. [?]</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3001.61</td>
<td>1</td>
<td>3003.56</td>
<td>3005.52</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>case4gs</td>
<td>Grainger and Stevenson [?]</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>10</td>
<td>2316.37</td>
<td>1</td>
<td>3681.07</td>
<td>4595.8</td>
<td>0.37</td>
<td></td>
</tr>
<tr>
<td>case5w</td>
<td>Lesieutre et al. [?]</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>2003.57</td>
<td>1</td>
<td>2004.6</td>
<td>2005.63</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>case6ac</td>
<td>McCoy et al.</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>23</td>
<td>4005.44</td>
<td>1</td>
<td>5574.23</td>
<td>6013.26</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>case6ac2</td>
<td>McCoy et al.</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>22</td>
<td>4005.32</td>
<td>1</td>
<td>5554.03</td>
<td>6012.82</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>case6b</td>
<td>McCoy et al.</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>30</td>
<td>3005.7</td>
<td>5</td>
<td>3606.3</td>
<td>4508.28</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>case6cyc</td>
<td>Bukhsh et al. [?]</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>30</td>
<td>3005.7</td>
<td>4</td>
<td>3606.3</td>
<td>4508.28</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>case6cyc2</td>
<td>McCoy et al.</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td>1506.87</td>
<td>1</td>
<td>3131.55</td>
<td>4508.24</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>case6cyc3</td>
<td>McCoy et al.</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>12</td>
<td>4502.42</td>
<td>1</td>
<td>4505.95</td>
<td>4508.01</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>case7</td>
<td>McCoy et al.</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>1667.31</td>
<td>1</td>
<td>1668.36</td>
<td>1669.41</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>case8cyc</td>
<td>Bukhsh et al. [?]</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>60</td>
<td>6506.11</td>
<td>1</td>
<td>8557.72</td>
<td>9511.04</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>case9</td>
<td>Chow [?]</td>
<td>9</td>
<td>9</td>
<td>2</td>
<td>16</td>
<td>3504.44</td>
<td>1</td>
<td>5692.02</td>
<td>6505.02</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>case9g</td>
<td>McCoy et al.</td>
<td>9</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td>2604.41</td>
<td>1</td>
<td>4103.36</td>
<td>5602.31</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td>caseK4</td>
<td>McCoy et al.</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>2008.62</td>
<td>1</td>
<td>3006.32</td>
<td>4005.51</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>caseK4sym</td>
<td>McCoy et al.</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>6</td>
<td>2009.28</td>
<td>2</td>
<td>3200.00</td>
<td>4005.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>caseK6</td>
<td>McCoy et al.</td>
<td>6</td>
<td>15</td>
<td>5</td>
<td>36</td>
<td>2018.2</td>
<td>1</td>
<td>2020.0</td>
<td>4005.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>caseK6</td>
<td>McCoy et al.</td>
<td>6</td>
<td>15</td>
<td>5</td>
<td>48</td>
<td>6002.79</td>
<td>1</td>
<td>6003.01</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>
Corollary

In the alternating-current optimal power flow problem, i.e., with \( s > 1 \), where powers are variable outside of the reference bus and there are no additional inequalities, the complex solution set is empty or positive-dimensional, except for a parameter set of measure zero. When the complex solution set is positive-dimensional, if a smooth real feasible solution exists, then there are infinitely many real feasible solutions.
The Agenda

An Introduction
Some structural results

**Convexifications based on semidefinite programming (SDP)**
SDP solvers based on tree-width decompositions, $r$-space sparsity, and interior-point methods
SDP solvers based on coordinate descent
Solvers combining a convexification with Newton method on the non-convex problem
The SDP Relaxation of Lavaei and Low

Lavaei and Low cast the R-PQV ACOPF without the extensions as:

\[
\min_{V,P^G,Q^G} \sum_{l=1}^{m} f_l(P_l) \tag{11}
\]

\[
P_{\text{min}}^{} \leq P^G \leq P_{\text{max}}^{} \tag{12}
\]

\[
Q_{\text{min}}^{} \leq Q^G \leq Q_{\text{max}}^{} \tag{13}
\]

\[
V_{k_{\text{min}}}^{} \leq |V_k| \leq V_{k_{\text{max}}}^{} \tag{14}
\]

\[
\Re\{ V_i(V_i - V_m)^* y_{lm}^* \} \leq P_{lm_{\text{max}}}^{} \tag{15}
\]

\[
\text{trace}\{ WY^* e_k^* e_k^* \} = P_k + Q_k \tag{16}
\]

\[
W \succeq 0, \ W = V V^* \tag{17}
\]

\[
\text{rank}(W) = 1 \tag{18}
\]
The SDP Relaxation of Lavaei and Low

Drop the rank constraint to obtain a SDP:

$$\min_{X, P^G, Q^G} \sum_{l=1}^{m} [c_l 2 (\text{trace}\{Y_k W\} + P_{D_k})^2 + c_l 1 \text{trace}\{Y_k W\} + P_{D_k}]$$

(19)

$$P^{\text{min}} \leq \text{trace}\{Y_k W\} \leq P^{\text{max}}$$

(20)

$$Q^{\text{min}} \leq \text{trace}\{\bar{Y}_k W\} \leq Q^{\text{max}}$$

(21)

$$(V_k^{\text{min}})^2 \leq \text{trace}\{M_k W\} \leq (V_k^{\text{max}})^2$$

(22)

$$\text{trace}\{Y_{lm} W\}^2 + \text{trace}\{\bar{Y}_{lm} W\}^2 \leq (S_{lm}^{\text{max}})^2$$

(23)

$$\text{trace}\{Y_{lm} W\} \leq P_{lm}^{\text{max}}$$

(24)

$$\text{trace}\{M_{lm} W\} \leq (V_{lm}^{\text{max}})^2$$

(25)

$$W = XX^T$$

(26)

There is no duality gap under some spectral conditions.
How to See the SDP?

For a polynomial optimisation problem with multi-variate polynomials $f, g_i$:

$$\min f(x) \text{s.t. } g_i(x) \geq 0 \quad i = \{1, \ldots, m\} \quad \text{[PP]}$$

one use the ideas of Shor (1987) and Nesterov (2000) to construct:

- The DSOS/SDSOS hierarchy of Ahmadi and Majumdar (2014)
- The SOCP hierarchy of Kuang et al (2015)
- The variant hierarchy of Ma et al (2016)

The first levels of the hierarchies of Lasserre, Waki, Ghaddar, and Ma are equivalent to the dual of the SDP of previous slide.
The Lasserre Hierarchy

A Conic Reformulation

Given $S \subseteq \mathbb{R}^n$, define $P_d(S)$ to be the cone of polynomials of degree at most $d$ that are non-negative over $S$

$$\max \varphi \quad \text{s.t.} \quad f(x) - \varphi \geq 0 \quad \forall \ x \in S_G,$$

$$= \max \varphi \quad \text{s.t.} \quad f(x) - \varphi \in P_d(S_G).$$

for $G = \{g_i(x) : i = 1, \ldots, m\}$, $S_G = \{x \in \mathbb{R}^n : g(x) \geq 0, \ \forall g \in G\}$.

The hierarchy of Lasserre approximates $P_d(S_G)$ by sums-of-squares

$$K_r^G = \Sigma_r + \sum_{i=1}^{m} g_i(x) \Sigma_{r - \deg(g_i)},$$

where $\Sigma_d := \{\sum_{i=1}^{N} p_i(x)^2 : p_i(x) \in \mathbb{R}_{\lfloor \frac{d}{2} \rfloor}[x]\}$, with $N = \binom{n+d}{d}$ for $r > d$. 
The Lasserre Hierarchy II

The Corresponding Optimisation Problem \([PP-Hr]\)

\[
\begin{align*}
\max_{\varphi, \sigma_i(x)} & \quad \varphi \quad \text{s.t.} \quad f(x) - \varphi = \sigma_0(x) + \sum_{i=1}^{m} \sigma_i(x)g_i(x) \\
& \quad \sigma_0(x) \in \Sigma_r, \\
& \quad \sigma_i(x) \in \Sigma_{r - \text{deg}(g_i)}. 
\end{align*}
\]

Notice that this is a semidefinite programming (SDP) problem, i.e.

\[
\min_{X \in \mathbb{S}^n} \langle C, X \rangle_{\mathbb{S}^n} \quad \text{s.t.} \quad \langle A_i, X \rangle_{\mathbb{S}^n} = b_i, \quad i = 1, \ldots, m, \quad X \succeq 0
\]
The Code using YALMIP

One can define a bi-variate polynomial using, e.g.

```latex
sdpvar x y
p = (1 + x)^4 + (1 - y)^2; Then:
```

### An implementation of [PP-H1]

```latex
v = monolist([x y],degree(p)/2);
Q = sdpvar(length(v));
sos = v'*Q*v;
F = [coefficients(p-sos,[x y]) == 0, Q >= 0];
optimize(F)
```
The Convergence

Summarising the results of Lasserre and others:

**Proposition (arXiv:1404.3626)**

Under mild assumptions, for the dense hierarchy \([PP-H_r]*\) for \([PP]\) and the respective duals the following holds:

(a) \(\inf [PP-H_r] \uparrow \min ([PP]) \) as \(r \to \infty\),

(b) \(\sup [PP-H_r]^* \uparrow \min ([PP]) \) as \(r \to \infty\),

(c) If the interior of the feasible set of \([PP]\) is nonempty, there is no duality gap between \([PP-H_r]\) and \([PP-H_r]^*\).

(d) If \([PP]\) has a unique global minimizer, \(x^*\), then as \(r\) tends to infinity the components of the optimal solution of \([PP-H_r]\) corresponding to the linear terms converge to \(x^*\).

See arXiv:1404.3626 for a discussion of assumptions, which are satisfied in the following applications.
The Agenda

An Introduction
Some structural results
Convexifications based on semidefinite programming (SDP)
**SDP solvers based on tree-width decompositions, r-space sparsity, and interior-point methods**
SDP solvers based on coordinate descent
Solvers combining a convexification with Newton method on the non-convex problem
Motivational Eye Candy

For a 3 bus, 3 branch system:

\[ r=2 \]
Motivational Eye Candy

For a 3 bus, 3 branch system:

\[ r=4 \]
Motivational Eye Candy

For a 3 bus, 3 branch system:
Motivational Eye Candy

For a 3 bus, 3 branch system:
An Overview

Let us formulate a different hierarchy of SDPs using:

- **Pre-processing**: Drop some monomials, consider sign symmetry, etc.
- **Correlative sparsity pattern**: The underlying network and hence the canonical hypergraph have low tree-width.
- **Range-space sparsity**: Any matrix can be reordered to a block-diagonal form with a border; each block becomes an SDP constraint, plus equalities for the border.
- **Facial reduction**: You can pre-solve away the equalities to regain some constraint qualification.
Correlative Sparsity [Waki et al. 2006]

The $n \times n$ Correlative Sparsity Pattern Matrix

\[ R_{ij} = \begin{cases} 
\star & \text{for } i = j \\
\star & \text{for } x_i, x_j \text{ in the same monomial of } f \\
\star & \text{for } x_i, x_j \text{ in the same constraint } g_k \\
0 & \text{otherwise}, 
\end{cases} \]

The chordal extension of its adjacency graph $G$ have maximal cliques $\{I_k\}_{k=1}^p$, $I_k \subset \{1, \ldots, n\}$. The sparse approximation of $\mathcal{P}_d(S)$ is

\[ \mathcal{K}_G^r(I) = \sum_{k=1}^p \left( \Sigma_r(I_k) + \sum_{j \in J_k} g_j \Sigma_{r-\deg(g_j)}(I_k) \right), \]

where $\Sigma_d(I_k)$ is the set of all sum-of-squares polynomials of degree up to $d$ supported on $I_k$ and $(J_1, \ldots, J_p)$ is a partitioning of the set of polynomials $\{g_j\}_j$ defining $S$ such that for every $j$ in $J_k$, the corresponding $g_j$ is supported on $I_k$. The support $I \subset \{1, \ldots, n\}$ of a polynomial contains $i$ of terms $x_i$ in one of the monomials.
Correlative Sparsity II

Another Hierarchy of Relaxations ([PP-SH$_r$]*) [Waki et al. 2006]

\[
\begin{align*}
\max_{\varphi, \sigma_k(x), \sigma_{r,k}(x)} & \quad \varphi \\
\text{s.t.} \quad f(x) - \varphi &= \sum_{k=1}^{p} \left( \sigma_k(x) + \sum_{j \in J_k} g_j(x) \sigma_{j,k}(x) \right) \\
\sigma_k &\in \Sigma_r((I_k)), \sigma_{j,k} \in \Sigma_{r-\deg(g_j)}(I_k).
\end{align*}
\]

Reduction in size:

Size of $k$ matrix inequalities $\left( \begin{array}{c} |C_k| \\ +r \\ |C_k| \end{array} \right)$ vastly smaller if $|C_k| \ll n$. 
The Convergence

Summarising the results of Waki, Kojima and others:

**Proposition (arXiv:1404.3626)**

Under mild assumptions, for the sparse hierarchy \([PP-SH_r]^*\) for \([PP]\) and the respective duals the following holds:

(a) \(\inf [PP-SH_r] \to \min ([PP])\) as \(r \to \infty\),

(b) \(\sup [PP-SH_r]^* \to \min ([PP])\) as \(r \to \infty\),

(c) If the interior of the feasible set of \([PP]\) is nonempty, there is no duality gap between \([PP-SH_r]\) and \([PP-SH_r]^*\).

(d) If \([PP]\) has a unique global minimizer, \(x^*\), then as \(r\) tends to infinity the components of the optimal solution of \([PP-SH_r]\) corresponding to the linear terms converge to \(x^*\).
Sparsity in SDPs [Kim et al. 2011]

\[
\begin{aligned}
\min \ b^T y \ \text{s.t.} \quad & F(y) \succeq 0, \ y \in \mathbb{R}^n, \\
\text{where} \\
F(y) &= \begin{pmatrix}
1 - y_1^4 & 0 & 0 & \cdots & 0 & y_1y_2 \\
0 & 1 - y_2^4 & 0 & \cdots & 0 & y_2y_3 \\
0 & 0 & \ddots & \cdots & 0 & y_3y_4 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & 0 & 1 - y_{n-1}^4 & y_{n-1}y_n \\
y_1y_2 & y_2y_3 & y_3y_4 & \cdots & y_{n-1}y_n & 1 - y_n^4
\end{pmatrix}.
\end{aligned}
\]

No correlative sparsity, but r-space sparsity!
Applying d-space and r-space conversion yields equivalent problem:

\[
\begin{align*}
\text{min} \quad & b^T y \\
\text{s.t.} \quad & \begin{pmatrix}
1 - y_1^4 & y_1 y_2 \\
y_1 y_2 & z_1 \\
1 - y_i^4 & y_i y_{i+1} \\
y_i y_{i+1} & -z_{i-1} + z_i \\
1 - y_{n-1}^4 & y_{n-1} y_n \\
y_{n-1} y_n & 1 - y_n^4 - z_{n-2}
\end{pmatrix} \succeq 0, \\
& \quad i = 2, 3, \ldots, n - 2, \\
& \quad (i = 2, 3, \ldots, n - 2), \\
& \quad i = 2, 3, \ldots, n - 2),
\end{align*}
\]

Equivalent problem satisfies correlative sparsity pattern!
**What can Off-The-Shelf Solvers Do**

Following much Matlab pre-processing, SeDuMi performs ok:

<table>
<thead>
<tr>
<th>Instance</th>
<th>Bound</th>
<th>Dim</th>
<th>Time</th>
<th>Bound</th>
<th>L2</th>
<th>Dim</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>case9mod</td>
<td>2753.23</td>
<td>588×168</td>
<td>0.6</td>
<td>3087.89</td>
<td>1792×14847</td>
<td>17.5</td>
<td></td>
</tr>
<tr>
<td>case14</td>
<td>8081.52</td>
<td>888×94</td>
<td>0.4</td>
<td>8081.54</td>
<td>7508×66740</td>
<td>1420.5</td>
<td></td>
</tr>
<tr>
<td>case14mod</td>
<td>7792.72</td>
<td>888×94</td>
<td>0.9</td>
<td>7991.07</td>
<td>7508×66740</td>
<td>904.2</td>
<td></td>
</tr>
<tr>
<td>case30mod</td>
<td>576.89</td>
<td>4706×684</td>
<td>3.8</td>
<td>578.56</td>
<td>36258×49164</td>
<td>13740.0</td>
<td></td>
</tr>
<tr>
<td>case39</td>
<td>41862.08</td>
<td>7282×758</td>
<td>2.2</td>
<td>41864.18</td>
<td>26076×215772</td>
<td>4359.1</td>
<td></td>
</tr>
<tr>
<td>case57</td>
<td>41737.79</td>
<td>13366×356</td>
<td>3.2</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>case118</td>
<td>129654.62</td>
<td>56620×816</td>
<td>6.1</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>case300</td>
<td>719711.63</td>
<td>362076×1938</td>
<td>13.6</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>case2383wp</td>
<td>1.814×10^6</td>
<td>22778705×47975</td>
<td>3731.5</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>case2736sp</td>
<td>1.307×10^6</td>
<td>30019740×57408</td>
<td>3502.2</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>case2746wp</td>
<td>1.584×10^6</td>
<td>30239940×57978</td>
<td>3903.2</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

What the Off-The-Shelf Solvers cannot Do

Unlike other general-purpose solvers (on L1):

<table>
<thead>
<tr>
<th>Instance</th>
<th>Time [s]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SDP</td>
</tr>
<tr>
<td>case2w</td>
<td>P</td>
</tr>
<tr>
<td>case2w</td>
<td>D</td>
</tr>
<tr>
<td>case5w</td>
<td>P</td>
</tr>
<tr>
<td>case5w</td>
<td>D</td>
</tr>
<tr>
<td>case9</td>
<td>P</td>
</tr>
<tr>
<td>case9</td>
<td>D</td>
</tr>
<tr>
<td>case14</td>
<td>P</td>
</tr>
<tr>
<td>case14</td>
<td>D</td>
</tr>
<tr>
<td>case30</td>
<td>P</td>
</tr>
<tr>
<td>case30</td>
<td>D</td>
</tr>
<tr>
<td>case30Q</td>
<td>P</td>
</tr>
<tr>
<td>case30Q</td>
<td>D</td>
</tr>
<tr>
<td>case39</td>
<td>P</td>
</tr>
<tr>
<td>case39</td>
<td>D</td>
</tr>
<tr>
<td>case57</td>
<td>P</td>
</tr>
<tr>
<td>case57</td>
<td>D</td>
</tr>
<tr>
<td>case118</td>
<td>P</td>
</tr>
<tr>
<td>case118</td>
<td>D</td>
</tr>
<tr>
<td>case300</td>
<td>P</td>
</tr>
<tr>
<td>case300</td>
<td>D</td>
</tr>
</tbody>
</table>
What the Off-The-Shelf Solvers cannot Do

- Run on a standard desktop
- “Good feasible solutions quick”
- Overall run-times comparable to MATPOWER, a popular heuristic
- Parallel and/or distributed computation for the SDP relaxations
- Exact tree-width decomposition fast – after all, it’s NP-Hard
- Exploitation of the structure in higher-level relaxations
- Warm start: a solution to [PP-H₁] is a very good solution to [PP-H₂], for an appropriate mapping of variables
The Agenda

An Introduction
Some structural results
Convexifications based on semidefinite programming (SDP)
SDP solvers based on tree-width decompositions, \( r \)-space sparsity, and interior-point methods

**SDP solvers based on coordinate descent**
Solvers combining a convexification with Newton method on the non-convex problem
Motivational Eye Candy

The graph shows the relationship between the number of threads and speedup for various cases. The cases are labeled as follows:
- case2224
- case118mod
- case9241pegase
- case2736sp
- case3w

The y-axis represents speedup, and the x-axis represents the number of threads. The graph demonstrates how speedup increases as the number of threads increases.
The Overview

- Trivial per-iteration run-time and memory requirements, thanks to a closed-form step for both ACOPF and general POP
- Warm start capabilities, due to the use of first-order methods
- Parallel and distributed computation easy to implement
- Overall run-times comparable to MATPOWER, a popular heuristic
Another Reformulation

\[
\min \ F(W) := \sum_{k \in G} f_k(W) \quad \quad \quad \quad \quad \quad \quad \quad [\text{RrBC}]
\]

s.t. \quad t_k = \text{tr}(Y_k W)

(27)

\[
P_k^{\min} - P_k^{\max} \leq t_k \leq P_k^{\max} - P_k^{\min}
\]

(28)

\[
g_k = \text{tr}({\bar{Y}}_k W)
\]

(29)

\[
Q_k^{\min} - g_k \leq g_k \leq Q_k^{\max} - g_k
\]

(30)

\[
h_k = \text{tr}(M_k W)
\]

(31)

\[
(V_k^{\min})^2 \leq h_k \leq (V_k^{\max})^2
\]

(32)

\[
u_{lm} = \text{tr}(Y_{lm} W)
\]

(33)

\[
u_{lm} = \text{tr}({\bar{Y}}_{lm} W)
\]

(34)

\[
z_{lm} = (u_{lm})^2 + (v_{lm})^2
\]

(35)

\[
z_{lm} \leq (S_{lm}^{\max})^2
\]

(36)

\[
W \succeq 0, \quad \text{rank}(W) \leq r.
\]
Some Observations

$O_1$: constraints (27–34) and (36) are box constraints,

$O_2$: using elementary liner algebra:

$$
\{ W \in S^n \text{ s.t. } W \succeq 0, \text{ rank}(W) \leq r \} = \{ RR^T \text{ s.t. } R \in \mathbb{R}^{n \times r} \},
$$

$O_3$: if rank($W^*$) > 1 for the optimum $W^*$ of the Lavaei-Low relaxation, there are no known methods for extracting the global optimum of [R1BC] from $W$, except for going higher in the hierarchies.

$O_4$: zero duality gap at any SDP relaxation in the hierarchy of Lasserre does not guarantee the solution of the SDP relaxation is exact for [R1BC].
Solving the SDP Relaxation

- Use the low-rank method of Burer and Monteiro, i.e. for increasing rank $r$, the augmented Lagrangian in variable $RR^T \in \mathbb{R}^{n \times n}$ rather than $W \in \mathbb{R}^{n \times n}$:

$$
\mathcal{L}(R, t, h, g, u, v, z, \lambda^t, \lambda^g, \lambda^h, \lambda^u, \lambda^v, \lambda^z) := \sum_{k \in G} f_k(RR^T)
- \sum_k \lambda_k^t (t_k - \text{trace}(Y_k RR^T)) + \frac{\mu}{2} \sum_k (t_k - \text{trace}(Y_k RR^T))^2
- \sum_k \lambda_k^g (g_k - \text{trace}(\bar{Y}_k RR^T)) + \frac{\mu}{2} \sum_k (g_k - \text{trace}(\bar{Y}_k RR^T))^2
- \sum_k \lambda_k^h (h_k - \text{trace}(M_k RR^T)) + \frac{\mu}{2} \sum_k (h_k - \text{trace}(M_k RR^T))^2
- \sum_{(l,m)} \lambda_{(l,m)}^u (u(l,m) - \text{trace}(Y(l,m) RR^T)) + \frac{\mu}{2} \sum_{(l,m)} (u(l,m) - \text{trace}(Y(l,m) RR^T))^2
- \sum_{(l,m)} \lambda_{(l,m)}^v (v(l,m) - \text{trace}(\bar{Y}(l,m) RR^T)) + \frac{\mu}{2} \sum_{(l,m)} (v(l,m) - \text{trace}(\bar{Y}(l,m) RR^T))^2
- \sum_{(l,m)} \lambda_{(l,m)}^z (z(l,m) - u^2(l,m) - v^2(l,m)) + \frac{\mu}{2} \sum_{(l,m)} (z(l,m) - u^2(l,m) - v^2(l,m))^2 + \nu \mathcal{R}.
$$

- Combined with a parallel coordinate descent with a closed-form step
Solving the SDP Relaxation II

1: for \( r = 1, 2, \ldots \) do
2: \hspace{1em} choose \( R \in \mathbb{R}^{m \times r} \)
3: \hspace{1em} compute corresponding values of \( t, h, g, u, v, z \)
4: \hspace{1em} project \( t, h, g, u, v, z \) onto the box constraints
5: for \( k = 0, 1, 2, \ldots \) do
6: \hspace{2em} in parallel, minimize \( \mathcal{L} \) in \( t, g, h, u, v, z \), coordinate-wise,
7: \hspace{2em} in parallel, minimize \( \mathcal{L} \) in \( R \), coordinate-wise
8: \hspace{1em} update Lagrange multipliers \( \lambda^t, \lambda^g, \lambda^h, \lambda^u, \lambda^v, \lambda^z \)
9: \hspace{1em} update \( \mu \)
10: \hspace{1em} terminate, if criteria are met
11: end for
12: end for
An Illustration

[Graph showing iterations vs. objective value with various markers and labels for different values such as 1e-01, 1e-02, 1e-03, 1e-04, 1e-05, 1e-06, and adaptive.]
An Illustration
An Illustration

![Graph showing speedup vs number of threads for different cases. The graph has lines for case 224, case 118mod, case 9241pegase, and case 2736sp. Each line represents the speedup at various numbers of threads.]
Let’s compare the run-times with:

- **MATPOWER**, a popular heuristic
- **sdp_pf**, the pre-processor of Molzahn et al. with SeDuMi
- **OPF_Solver**, the pre-processor of Lavaei et al. with SDPT3

<table>
<thead>
<tr>
<th>Name</th>
<th>MATPOWER</th>
<th>sdp_pf</th>
<th>OPF_Solver</th>
<th>Our Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>case6ww</td>
<td>3.144e+03</td>
<td>0.114</td>
<td>3.144e+03</td>
<td>0.74</td>
</tr>
<tr>
<td>case14</td>
<td>8.082e+03</td>
<td>0.201</td>
<td>8.082e+03</td>
<td>0.84</td>
</tr>
<tr>
<td>case30</td>
<td>5.769e+02</td>
<td>0.788</td>
<td>5.769e+02</td>
<td>2.70</td>
</tr>
<tr>
<td>case39</td>
<td>4.189e+04</td>
<td>0.399</td>
<td>4.189e+04</td>
<td>3.26</td>
</tr>
<tr>
<td>case57</td>
<td>4.174e+04</td>
<td>0.674</td>
<td>4.174e+04</td>
<td>2.69</td>
</tr>
<tr>
<td>case118</td>
<td>1.297e+05</td>
<td>1.665</td>
<td>1.297e+05</td>
<td>6.57</td>
</tr>
<tr>
<td>case300</td>
<td>7.197e+05</td>
<td>2.410</td>
<td>–</td>
<td>17.68</td>
</tr>
</tbody>
</table>
# What can this Do

Let's look at instances from NESTA:

| Instance | MATPOWER |  | sdp_pf |  | Our Approach |  |
|----------|----------|---------------|--------|---------------|----------|
| NESTA    |          |               |        |               |          |         |
| case3_lmbd | 5.812e+03 | 0.946         | 5.789e+03 | 2.254         | 5.757e+03 | 0.149   |
| case4_gs  | 1.564e+02 | 1.019         | 1.564e+02 | 2.392         | 1.564e+02 | 0.139   |
| case5_pjm | (1.599e-01) | 0.811        | (1.599e-01) | 2.708        | 2.008e+04 | 0.216   |
| case6_c   | 2.320e+01 | 0.825         | 2.320e+01 | 2.392         | 2.320e+01 | 0.379   |
| case6_ww  | 3.143e+03 | 0.884         | 3.143e+03 | 2.776         | 3.148e+03 | 0.242   |
| case9_wsc | 5.296e+03 | 1.077         | 5.296e+03 | 2.621         | 5.296e+03 | 0.211   |
| case14_ieee | 2.440e+02 | 0.762        | 2.440e+02 | 3.164         | 2.440e+02 | 0.267   |
| case30_as | 8.031e+02 | 0.861         | 8.031e+02 | 3.916         | 8.031e+02 | 0.608   |
| case30_fsr | 5.757e+02 | 0.898         | 5.757e+02 | 6.201         | 5.750e+02 | 0.613   |
| case30_ieee | 2.049e+02 | 0.818         | 2.049e+02 | 4.335         | 2.049e+02 | 0.788   |
| case39_epri | 9.651e+04 | 0.863        | 9.649e+04 | 5.676         | 9.651e+04 | 0.830   |
| case57_ieee | 1.143e+03 | 1.119         | 1.143e+03 | 8.053         | 1.143e+03 | 0.957   |
| case118_ieee | (4.433e+02) | 1.181       | (4.433e+02) | 18.097       | 4.098e+03 | 4.032   |
| case162_dtc | (3.123e+03) | 0.975       | (3.123e+03) | 32.153       | 4.215e+03 | 8.328   |
## Let’s look at API instances from NESTA:

<table>
<thead>
<tr>
<th>Instance</th>
<th>MATPOWER</th>
<th>sdppf</th>
<th>Our Approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>case3_lmbd</td>
<td>3.677e+02</td>
<td>1.113</td>
<td>3.333e+02</td>
</tr>
<tr>
<td>case5_pjm</td>
<td>(5.953e+00)</td>
<td>1.002</td>
<td>(5.953e+00)</td>
</tr>
<tr>
<td>case6_c</td>
<td>8.144e+02</td>
<td>1.034</td>
<td>8.144e+02</td>
</tr>
<tr>
<td>case9_wscc</td>
<td>6.565e+02</td>
<td>1.135</td>
<td>6.565e+02</td>
</tr>
<tr>
<td>case14_ieee</td>
<td>3.255e+02</td>
<td>1.035</td>
<td>3.255e+02</td>
</tr>
<tr>
<td>case30_as</td>
<td>5.711e+02</td>
<td>1.156</td>
<td>5.711e+02</td>
</tr>
<tr>
<td>case30_fsr</td>
<td>3.721e+02</td>
<td>0.862</td>
<td>3.309e+02</td>
</tr>
<tr>
<td>case30_ieee</td>
<td>4.155e+02</td>
<td>1.076</td>
<td>4.155e+02</td>
</tr>
<tr>
<td>case39_epri</td>
<td>(1.540e+02)</td>
<td>0.965</td>
<td>(1.540e+02)</td>
</tr>
<tr>
<td>case57_ieee</td>
<td>1.430e+03</td>
<td>1.041</td>
<td>1.429e+03</td>
</tr>
<tr>
<td>case162_dtc</td>
<td>(1.502e+03)</td>
<td>1.215</td>
<td>(1.502e+03)</td>
</tr>
</tbody>
</table>
What can this Do

Let’s recall the performance of general-purpose solvers:

<table>
<thead>
<tr>
<th>Instance</th>
<th>SDP</th>
<th>SeDuMi 1.32</th>
<th>SDPA 7.0</th>
<th>SDPLR 1.03</th>
<th>CSDP 6.0.1</th>
<th>SDPT3 4.0</th>
<th>Mosek 7.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>case2w</td>
<td>P</td>
<td>0.3</td>
<td>0.003</td>
<td>0</td>
<td>0.02</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>case2w</td>
<td>D</td>
<td>0.3</td>
<td>0.004</td>
<td>0</td>
<td>0.02</td>
<td>0.4</td>
<td>0.1</td>
</tr>
<tr>
<td>case5w</td>
<td>P</td>
<td>0.4</td>
<td>–</td>
<td>0</td>
<td>0.04</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>case5w</td>
<td>D</td>
<td>0.4</td>
<td>–</td>
<td>0</td>
<td>0.03</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>case9</td>
<td>P</td>
<td>1.0</td>
<td>0.05</td>
<td>25</td>
<td>0.1</td>
<td>0.5</td>
<td>0.3</td>
</tr>
<tr>
<td>case9</td>
<td>D</td>
<td>1.0</td>
<td>0.07</td>
<td>25</td>
<td>0.1</td>
<td>0.6</td>
<td>0.3</td>
</tr>
<tr>
<td>case14</td>
<td>P</td>
<td>0.7</td>
<td>0.05</td>
<td>58</td>
<td>0.1</td>
<td>–</td>
<td>0.2</td>
</tr>
<tr>
<td>case14</td>
<td>D</td>
<td>0.7</td>
<td>0.04</td>
<td>17</td>
<td>0.1</td>
<td>–</td>
<td>0.3</td>
</tr>
<tr>
<td>case30</td>
<td>P</td>
<td>2.8</td>
<td>–</td>
<td>829</td>
<td>0.8</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>case30</td>
<td>D</td>
<td>6.1</td>
<td>–</td>
<td>282</td>
<td>1.0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>case30Q</td>
<td>P</td>
<td>2.8</td>
<td>–</td>
<td>831</td>
<td>0.8</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>case30Q</td>
<td>D</td>
<td>6.2</td>
<td>–</td>
<td>195</td>
<td>1.0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>case39</td>
<td>P</td>
<td>4.4</td>
<td>–</td>
<td>2769</td>
<td>1.0</td>
<td>–</td>
<td>0.7</td>
</tr>
<tr>
<td>case39</td>
<td>D</td>
<td>7.7</td>
<td>–</td>
<td>723</td>
<td>1.3</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>case57</td>
<td>P</td>
<td>3.2</td>
<td>–</td>
<td>1930</td>
<td>0.5</td>
<td>–</td>
<td>0.7</td>
</tr>
<tr>
<td>case57</td>
<td>D</td>
<td>4.0</td>
<td>–</td>
<td>1175</td>
<td>1.0</td>
<td>–</td>
<td>1.8</td>
</tr>
<tr>
<td>case118</td>
<td>P</td>
<td>10.3</td>
<td>0.9</td>
<td>4400</td>
<td>3.4</td>
<td>–</td>
<td>1.7</td>
</tr>
<tr>
<td>case118</td>
<td>D</td>
<td>17.1</td>
<td>–</td>
<td>–</td>
<td>13.0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>case300</td>
<td>P</td>
<td>27.5</td>
<td>1.7</td>
<td>–</td>
<td>109.6</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>case300</td>
<td>D</td>
<td>133.7</td>
<td>–</td>
<td>–</td>
<td>66.7</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>
The Agenda

An Introduction
Some structural results
Convexifications based on semidefinite programming (SDP)
SDP solvers based on tree-width decompositions, $r$-space sparsity, and interior-point methods
SDP solvers based on coordinate descent
**Solvers combining a convexification with Newton method on the non-convex problem**
Motivation (case30)

Number of Epochs

Infeasibility

case30 extended

CD 2
CD 4
CD 8
CD 16
CD 32
CD 64
CD 128
CD 140

Number of Epochs

Infeasibility

0 200 400 600 800 1000

10^-25
10^-20
10^-15
10^-10
10^-5
10^0
10^5
Motivation (case30)

- Number of Epochs: 0, 200, 400, 600, 800, 1000
- Objective Value: 400, 600, 800, 1000, 1200, 1400

Graph showing objective value over number of epochs for case30 extended with different CD values (2, 4, 8, 16, 32, 64, 128, 140).
Motivation (case118)
Motivation (case2383)

Infeasibility vs. Number of Epochs

- Case2383wp extended
- CD 4
- CD 16
- CD 64
- CD 201

Graph shows the decrease in infeasibility with increasing number of epochs for different cases.
Motivation (case2383)

case2383wp extended

Number of Epochs

Objective Value

CD 4
CD 16
CD 64
CD 201

IBM Research - Ireland
The quadratic convergence of Newton method is hard to beat.

Nevertheless, without applying a convexification, Newton method can converge to particularly poor local optima of a non-convex problem.

Point-estimation theory of Smale offers a test, based solely on derivative-at-the-current-iterate, for inclusion in a domain of monotonicity of a zero of a polynomial.

Thereby, we can combine the first-order methods for the (hierarchy of) convex Lagrangians with Newton method for (a hierarchy of) non-convex Lagrangians, while preserving convergence guarantees associated with the convexification.
Newton’s method

Let \( f : \mathbb{R}^n \to \mathbb{R}^n \) be a non-linear system, \( D_f : \mathbb{R}^n \to \mathbb{R}^{n \times n} \) its Jacobian.

\[
N_f(x) = \begin{cases} 
  x - D_f(x)^{-1}f(x) & \text{if } D_f(x) \text{ is invertible} \\
  x & \text{otherwise}
\end{cases}
\]

\[
N_f^k(x) = (N_f \circ \cdots \circ N_f)(x).
\]

By Ralf Pfeifer:
Newton’s method

*approximate solution, associated solution

A point $x_0 \in \mathbb{R}^n$ is an approximate solution of $f(x) = 0$ with associated solution $\xi$ if $f(\xi) = 0$ and

$$\|x_k - \xi\| \leq \left(\frac{1}{2}\right)^{2^k-1} \|x_0 - \xi\|$$

for $x_k = N_f^k(x_0)$.

An approximate solution will converge quadratically to its associated solution under the sequence of Newton iterations.
\( \alpha \)-theoretic quantities

Let \( f : \mathbb{R}^n \to \mathbb{R}^n \) be a nonlinear system, \( D_f : \mathbb{R}^n \to \mathbb{R}^{n \times n} \) its Jacobian.

- \( \gamma = \gamma(f, x) := \sup_{k \geq 2} \left\| \frac{D_f(x)^{-1}D_f^k(x)}{k!} \right\|^{\frac{1}{k-1}} \)
- \( \beta = \beta(f, x) := \| x - N_f(x) \| = \| D_f(x)^{-1}f(x) \| \)
- \( \alpha = \alpha(f, x) := \beta(f, x) \cdot \gamma(f, x) \)
Main results of $\alpha$-theory

Theorem (Blum, Cucker, Shub, Smale, 2012)

Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a polynomial system and $x, y \in \mathbb{R}^n$ with $x \neq y$.

- If $x$ is an approximate solution of $f = 0$ with associated solution $\xi$, then $\|x - \xi\| \leq 2\beta(f, x)$.
- If $\alpha(f, x) < \frac{13 - 3\sqrt{17}}{4} \approx 0.157671$, then $x$ is an approximate solution of $f = 0$.
- If $\alpha(f, x) < 0.03$ with $\|x - y\| \cdot \gamma(f, x) < 0.05$, then $x$ and $y$ are approximate solutions of $f = 0$ with the same associated solution.
**α-theoretic quantities**

- \( \gamma = \gamma(f, x) := \sup_{k \geq 2} \left\| \frac{D_f(x)^{-1}D^k_f(x)}{k!} \right\|^{\frac{1}{k-1}} \)
- \( \beta = \beta(f, x) := \|x - N_f(x)\| = \|D_f(x)^{-1}f(x)\| \)
- \( \alpha = \alpha(f, x) := \beta(f, x) \cdot \gamma(f, x) \)

\( D^k_f(x) \) is the symmetric tensor whose entries are the order-\( k \) partial derivatives of \( f \) at \( x \). The norm is a norm on operators from the \( k \)-fold symmetric power of \( \mathbb{R}^n \) to \( \mathbb{R}^n \).

**This norm is difficult to compute.**
Bounding $\gamma(f, x)$

Define some auxiliary quantities:

- $\|x\|_1^2 := 1 + \sum_{i=1}^{n} |x_i|^2$
- $\Delta(d)(x)_{i,i} = d_{i}^{1/2} \|x\|_1^{d_{i}-1}$
- $\|g\|_2^2 = \sum_{|\nu| \leq d} |g_{\nu}|^2 \left( \frac{\nu_1! \cdots \nu_n!(d-|\nu|)!}{d!} \right)$
  where $g_{\nu} \in \mathbb{R}$, $\nu \in \mathbb{N}^n$, $|\nu| = \sum_i \nu_i$
- $\|f\|_2^2 = \sum_{i=1}^{n} \|f_i\|^2$
- $\mu(f, x) = \max\{1, \|f\| \cdot \|D_f(x)^{-1}\Delta(d)(x)\|\}$

Theorem (Shub and Smale, 1993)

If $x \in \mathbb{R}^n$ such that $D_f(x)$ is invertible, then

$$\gamma(f, x) \leq \frac{\mu(x, f)D^{3/2}}{2\|x\|_1}.$$  \hfill (38)
Example

Minimizing the 2-variable Rosenbrock function

\[ f(x, y) = (1 - x)^2 + 100(y - x^2)^2 \quad \text{(here } (a, b) = (1, 100)) \]

\[ \nabla f(x, y) = \begin{bmatrix} -2(1 - x) - 400x(y - x^2) \\ 200(y - x^2) \end{bmatrix} = 0 \]

\[ (x, y) = (1.001, 0.999); \quad \| (x, y) - (a, a) \|_2 = \sqrt{2} \cdot 10^{-3} \]

\[ \gamma(f, (x, y)) \leq 3341.051306 \]

\[ \beta(f, (x, y)) \approx 0.001858 \]

\[ \alpha(f, (x, y)) \leq 6.208154 \]
Putting it all together

1: \textbf{for } k = 0, 1, 2, \ldots \textbf{ do}
2: \quad \text{Update } (x, \lambda) \text{ using coordinate descent for } \nabla \tilde{L} = 0
3: \quad \textbf{if } \alpha(\nabla L, x) \leq \alpha_0 \textbf{ then}
4: \quad \quad \textbf{S } \leftarrow (x, \lambda)
5: \quad \textbf{for } \ell = 0, 1, 2, \ldots \textbf{ do}
6: \quad \quad \text{Update } (x, \lambda) \text{ using Newton step on } \nabla L = 0
7: \quad \quad \textbf{if } \tilde{L}(x, \lambda) > \tilde{L}(S) \textbf{ then}
8: \quad \quad \quad (x, \lambda) \leftarrow S \quad \textbf{break}
9: \quad \textbf{end if}
10: \quad \quad \textbf{end for}
11: \quad \textbf{end if}
12: \textbf{end for}
13: \textbf{end if}
Illustrations (case57)

Number of Epochs

Infeasibility

0 100 200 300 400 500 600 700 800

10^-25
10^-20
10^-15
10^-10
10^-5
0
10^0
10^5

CD
Hybrid
Newton

Number of Epochs
Illustrations (case57)

The graph shows the objective value plotted against the number of epochs for different optimization methods. The methods compared are CD, Hybrid, and Newton. Each method has a distinct pattern and convergence rate. The CD method shows a steady decrease in objective value with an initial sharp drop. The Hybrid method also shows a drop but with more fluctuations. The Newton method has a slower initial decrease but stabilizes at a lower value compared to the other two methods.
Illustrations (case118)

The graph shows the infeasibility against the number of epochs for different methods:
- CD (blue stars)
- Hybrid (red diamonds)
- Newton (yellow crosses)

The x-axis represents the number of epochs, ranging from 0 to 1600, while the y-axis represents the infeasibility on a logarithmic scale, ranging from $10^{-25}$ to $10^5$.
Illustrations (case118)

Number of Epochs

Objective Value

Case 118

CD

Hybrid

Newton
Illustrations (case300)

The graph shows the infeasibility of different optimization algorithms over the number of epochs. The algorithms compared are CD, Hybrid, and Newton. The y-axis represents infeasibility on a logarithmic scale, ranging from $10^{-20}$ to $10^{10}$. The x-axis represents the number of epochs, ranging from 0 to 3.

- CD algorithm starts with the highest infeasibility and shows a steep decrease, approaching $10^{-5}$ by the end of 3 epochs.
- Hybrid algorithm also starts with high infeasibility but decreases at a slower rate, reaching $10^{-5}$ by 2.5 epochs.
- Newton algorithm starts with the lowest infeasibility and decreases steadily, reaching $10^{-15}$ by 2.5 epochs.

Overall, the Newton algorithm shows the best performance in reducing infeasibility over the number of epochs.
Illustrations (case300)
Illustrations (case300)

Solution Time

Objective Value

Case300

CD

Hybrid

Newton
Illustrations (case300)

Solution Time

Active Set

$10^0$

$10^{-1}$

$10^{-2}$

Solution Time
Illustrations (case30)

The image shows a graph with the following axes:
- Y-axis: Infeasibility (on a log scale from $10^{-25}$ to $10^0$)
- X-axis: Number of Epochs (from 0 to 1000)

The graph compares different algorithms:
- CD
- Hybrid
- Newton

Each algorithm is represented by a different marker:
- CD: Blue stars
- Hybrid: Orange diamonds
- Newton: Yellow crosses

The graph illustrates the decrease in infeasibility over the number of epochs for each algorithm, with CD showing the fastest decrease, followed by Hybrid, and then Newton.
Illustrations (case2383wp)

Number of Epochs # 10

Infeasibility

case2383wp extended

CD

Hybrid

Newton

Number of Epochs

Infeasibility

10^15

10^10

10^5

10^0

10^-5

10^-10

10^-15

10^-20

0 2 4 6 8 10
Illustrations (case2383wp)

Number of Epochs

Objective Value

case2383wp extended

CD
Hybrid
Newton
Illustrations (case2383wp)

Number of Epochs # 10
0 2 4 6 8 10
Active Set
$10^{-2}$
$10^{-1}$
$10^{0}$
case2383wp extended
CD
Hybrid
Newton
More Observations

Even for polynomial optimisation, in general, one may avoid the back-tracking:

- A global (non-convex, polynomial) Lagrangian
- Taylor expansion of the square root in a square-the-square-root absolute value
- A local (non-convex, polynomial) Lagrangian, where tight inequalities are taken as equalities
References

- arXiv:1412.8054: Power Flow as an Algebraic System
- arXiv:1510.06797: Alternative Linear and Second-order Cone Approximation Approaches for Polynomial Optimization
Work in Progress

- Kuramoto Model as Algebraic Systems
  (with Tianran Chen, Dhagash Mehta, and Matthew Niemerg)
- Hybrid methods for generic polynomial optimisation
  (with Martin Takac and Jie Liu)
- Rates of convergence in SDP hierarchies
  (with Martin Takac and Jie Liu)
- A variant of the hierarchy of Lasserre
  (with Martin Mevissen and Wann-Jiun Ma)
- Control of the related non-linear systems
  (with Bob Shorten)
- Mixed-integer second-order cone formulations
  (with Horace Cheng, Andreas Grothey)
- Mixed-criticality robustness
Conclusions

- Steady state problems in power systems are an excellent motivation for the study of polynomial optimisation.
- Some hierarchies of convexifications are very hard to solve, beyond the first level.
- Algorithm engineering matters.

- We would love to hear from you!
- Jakub Marecek: jakub.marecek@ie.ibm.com
Definitions

Let $n \geq 0$ be an integer and $f(z)$ be a system of $n$ polynomial equations in $z \in \mathbb{C}^n$ with support $(A_1, \ldots, A_n)$:

$$
\begin{align*}
  f_1(z) &= \sum_{\alpha \in A_1} f_{1\alpha} z_1^{\alpha_1} z_2^{\alpha_2} \cdots z_n^{\alpha_n} \\
  \vdots & \quad \vdots \\
  f_n(z) &= \sum_{\alpha \in A_n} f_{n\alpha} z_1^{\alpha_1} z_2^{\alpha_2} \cdots z_n^{\alpha_n},
\end{align*}
$$

where coefficients $f_{i\alpha}$ are non-zero complex numbers. It is well known that the polynomials define $n$ projective hypersurfaces in a projective space $\mathbb{CP}^n$. 
Bézout theorem states that either hypersurfaces intersect in an infinite set with some component of positive dimension, or the number of intersection points, counted with multiplicity, is equal to the product $d_1 \cdots d_n$, where $d_i$ is the degree of polynomial $i$. We call the product $d_1 \cdots d_n$ the usual Bézout number.
Multihomogeneous Structure

**Definition (Structure)**

Any partition of the index set \( \{1, \ldots, n\} \) into \( k \) sets \( I_1, \ldots, I_k \) defines a structure. There, \( Z_j = \{ z_i : i \in I_j \} \) is known as the group of variables for each set \( I_j \). The associated degree \( d_{ij} \) of a polynomial \( f_i \) with respect to group \( Z_j \) is

\[
d_{ij} \overset{\text{def}}{=} \max_{\alpha \in A_i} \sum_{l \in I_j} \alpha_l.
\]

(40)

We say that \( f_i \) has *multi-degree* \( (d_{i1}, \ldots, d_{in}) \).
Multihomogeneous Structure

- Whenever for some $j$, for all $i$, the same $d_{ij}$ is attained for all $\alpha \in A_i$, we call the system homogeneous in the group of variables $Z_j$.
- The projective space associated to the group of variables $Z_j$ in a structure has dimension

$$a_j \overset{\text{def}}{=} \begin{cases} |l_j| - 1 & \text{if the system is homogeneous in } Z_j, \text{ and} \\ |l_j| & \text{otherwise.} \end{cases}$$

(41)
Multi-homogeneous Bézout theorem

Definition (Multi-homogeneous Bézout Number)

Assuming \( n = \sum_{j=1}^{k} a_j \), the multi-homogeneous Bézout number \( \text{Béz}(A_1, \ldots, A_n; l_1, \ldots, l_k) \) is defined as the coefficient of the term \( \prod_{j=1}^{k} \zeta_j^{a_j} \), where \( a_j \) is the associated dimension, within the polynomial \( \prod_{i=1}^{n} \sum_{j=1}^{k} d_{ij} \zeta_j \), in variables \( \zeta_j, j = 1 \ldots k \) with coefficients \( d_{ij} \) are the associated degrees; that is \( (d_{11} \zeta_1 + d_{12} \zeta_2 + \ldots + d_{1k} \zeta_k) (d_{21} \zeta_1 + d_{22} \zeta_2 + \ldots + d_{2k} \zeta_k) \cdots (d_{n1} \zeta_1 + d_{n2} \zeta_2 + \ldots + d_{nk} \zeta_k) \).
Multi-homogeneous Bézout theorem

Consider the example of Wampler in $x \in \mathbb{C}^3$:

$$\begin{align*}
p_1(z) &= x_1^2 + x_2 + 1, \\
p_2(z) &= x_1 x_3 + x_2 + 2, \\
p_3(z) &= x_2 x_3 + x_3 + 3,
\end{align*}$$

with the usual Bézout number of 8. Considering the partition $\{x_1, x_2\}$, $\{x_3\}$, where $d_{11} = 2$, $d_{12} = 0$, $d_{21} = d_{22} = d_{31} = d_{32} = 1$. the monomial $\zeta_1^2 \zeta_2^1$ is to be looked up in the polynomial $2\zeta_1 (\zeta_1 + \zeta_2)^2$. The corresponding multi-homogeneous Bézout number is hence 4 and this is the minimum across all possible structures.
Multi-homogeneous Bézout theorem

- In general, the multi-homogeneous Bézout number $\text{Béz}(A_1, \ldots, A_n; l_1, \ldots, l_k)$ is an upper bound on the number of isolated roots in $\mathbb{C}P^{a_1} \times \cdots \times \mathbb{C}P^{a_k}$, and thereby an upper bounds the number of isolated finite complex roots.

- There are a variety of additional methods for computing the multi-homogeneous Bézout number.

- In the particular case where $A = A_1 = \cdots = A_n$, we denote

$$\text{Béz}(A_1, \ldots, A_n; l_1, \ldots, l_k) \overset{\text{def}}{=} \left( \begin{array}{c} n \\ a_1 \ a_2 \ \cdots \ \ a_k \end{array} \right) \prod_{j=1}^{k} d_j^{a_j},$$

where $d_j = d_{ij}$ (equal for each $i$) and the multinomial coefficient

$$\left( \begin{array}{c} n \\ a_1 \ a_2 \ \cdots \ \ a_k \end{array} \right) \overset{\text{def}}{=} \frac{n!}{a_1! \ a_2! \ \cdots \ a_k!}$$

is the coefficient of $\prod_{j=1}^{k} \zeta_j^{a_k}$ in $(\zeta_1 + \cdots + \zeta_k)^n$ with $n = \sum_{i=1}^{k} a_i$, as above.
Multi-homogeneous Bézout theorem

- The multi-homogeneous Bézout number provides a sharper bound on the number of isolated solutions of a system of equations than the usual Bézout number $\prod_{i=1}^{n} d_i = d_1 \cdots d_n$.
- The example of the eigenvalue problem: the Bézout number is $2^n$ vs. structure with multi-homogeneous Bézout number of $n$. 
A bound on the number of isolated nonzero complex solutions which takes into consideration the monomials that appear in the polynomial system via Newton polytopes.

A *convex set* is a set of points in which the line segment connecting any pair of points in the set also lie in that set.

The *convex hull* of a set is the minimal convex set containing that set.

Given a polynomial, each of its terms give rise to an *exponent vector*. For instance, for the term $x^3y^2z^1$, the exponent vector is simply the vector whose entries are the exponents of $x$, $y$ and $z$, respectively, i.e., $(3, 2, 1)$. The choice of this ordering is inconsequential as long as it is kept the same for each equation.
• The set of all exponent vectors derived from the nonzero terms of an polynomial equation is called the support of that equation.
• For a polynomial, the convex hull of its support is known as the Newton polytope of that polynomial.
• Given \( n \) convex polytopes \( Q_1, \ldots, Q_n \subset \mathbb{R}^n \) and positive real numbers \( \lambda_1, \ldots, \lambda_n \) Minkowski’s Theorem states that the \( n \)-dimensional volume of the Minkowski sum \( \lambda_1 Q_1 + \cdots + \lambda_n Q_n \), defined as

\[
\left\{ \lambda_1 q_1 + \cdots + \lambda_n q_n \mid q_i \in Q_i \text{ for } i = 1, \ldots, n \right\}
\]

is a homogeneous polynomial of degree \( n \) in the variables \( \lambda_1, \ldots, \lambda_n \).
The coefficient associated with the monomial $\lambda_1 \cdots \lambda_n$ in this polynomial is known as the *mixed volume* of the polytopes $Q_1, \ldots, Q_n$.

In the simplest case, the mixed volume of two line segments on the plane is precisely the area of the parallelogram spanned by translations of these two line segments.