Conditional Moment Relaxations and Sums-of-AM/GM-Exponentials

Riley Murray

California Institute of Technology

MIT Virtual Seminar on Optimization and Related Areas
17 April 2020

The functions of interest

Caltech

polynomials

Parameters $\boldsymbol{a}_i$ in $\mathbb{N}^n$, c_i in $\mathbb{R}$.

Using $\boldsymbol{x}^{\boldsymbol{a}_i} = \prod_{j=1}^n x_j^{a_{ij}}$,

$$\boldsymbol{x} \mapsto \sum_{i=1}^m c_i \boldsymbol{x}^{\boldsymbol{a}_i}.$$

Care about degree: $\max_i \|\boldsymbol{a}_i\|_1$.

signomials

Parameters $\boldsymbol{a}_i$ in $\mathbb{R}^n$, c_i in $\mathbb{R}$.

In “exponential form”,

$$\boldsymbol{x} \mapsto \sum_{i=1}^m c_i \exp(\boldsymbol{a}_i \cdot \boldsymbol{x}).$$

Care about number of terms: m.

For historical and modeling reasons, signomials are often written in *geometric form*

$$\boldsymbol{y} \mapsto \sum_{i=1}^m c_i \boldsymbol{y}^{\boldsymbol{a}_i}$$

where $\boldsymbol{y} \in \mathbb{R}^n_{++}$ has the correspondence $y_i = \exp(x_i)$. **We use the exponential form!**

Geometric Programming

The signomial

$$f(\boldsymbol{x}) = \textstyle\sum_{i=1}^{m} c_i \exp(\boldsymbol{a}_i \cdot \boldsymbol{x})$$

is called a *posynomial* when all $c_i \geq 0$.

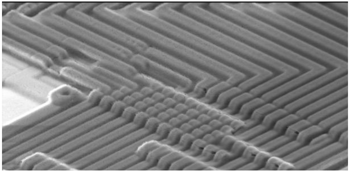
Geometric programs (GPs):

$$\inf_{\boldsymbol{x} \in \mathbb{R}^n} \{ f(\boldsymbol{x}) \, : \, g_i(\boldsymbol{x}) \leq 1 \, \forall \, i \in [k] \}$$

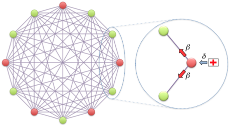
where f and $\{g_i\}_{i=1}^{k}$ are posynomials.

Study of GPs initiated by Zener, Duffin, and Peterson (1967). Exponential-form GPs are convex & poly-time solvable via IPMs [1].

Optimization-based engineering design: electrical [2, 3, 4], structural [5, 6], environmental [7], and aeronautical [8, 9].



Epidemilogical process control [10, 11, 12], power control and storage [13, 14], self-driving cars [15], gas network operation [16].



Additional applications in healthcare [17], biology [18], economics [19, 20, 21], and statistics [22, 23]

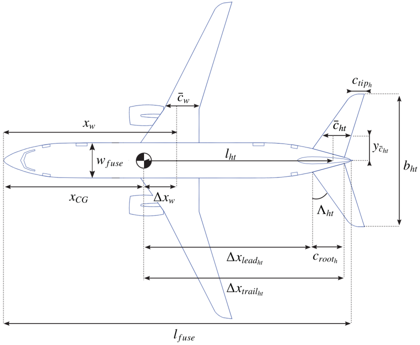
Signomial programming

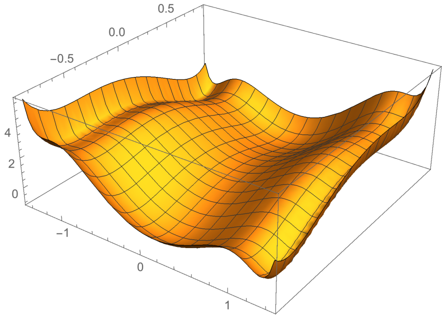
Caltech

A *signomial program* (SP) is an optimization problem stated with signomials, e.g.

$$\inf_{\boldsymbol{x} \in \mathbb{R}^n} \{ f(\boldsymbol{x}) \, : \, g_i(\boldsymbol{x}) \leq 0 \text{ for all } i \text{ in } [k] \}.$$

Major applications in aircraft design [24, 25, 26, 27, 28] and structrual engineering [29, 30, 31, 32]. Additional applications in EE [33], communications [34], and ML [35].





The AM/GM-inequality

Caltech

If $\boldsymbol{u}, \boldsymbol{\lambda} \in \mathbb{R}^m$ are positive and $\mathbf{1}^\intercal \boldsymbol{\lambda} = 1$, then

$$\boldsymbol{u}^{\boldsymbol{\lambda}} \leq \boldsymbol{\lambda}^\intercal \boldsymbol{u}.$$

Proof. If $\boldsymbol{v} = \log \boldsymbol{u}$, then $\boldsymbol{u}^{\boldsymbol{\lambda}} = \exp(\boldsymbol{\lambda}^\intercal \boldsymbol{v}) \leq \sum_{i=1}^m \lambda_i \exp v_i = \boldsymbol{\lambda}^\intercal \boldsymbol{u}$. $\square$

A recent history of using the AM/GM inequality to certify function nonnegativity:

- 1978 and 1989: Reznick [36, 37].
- 2009: Pébay, Rojas and Thompson [38].
- 2012 and 2013: Ghasemi and Marshall [39], Ghasemi, Lasserre, and Marshall [40].
- 2012: Paneta, Koeppl, and Craciun [41], and August, Craciun, and Koeppl [42].
- 2016: Iliman and de Wolff [43].

When used for computation, exponents $\{\boldsymbol{a}_i\}_{i=1}^m$ were presumed to be *highly structured*.

E.g. $\operatorname{conv}\{\boldsymbol{a}_i\}_{i=1}^m$ has $m-1$ extreme points, 1 point in its relative interior.

Definitions from convex analysis

A set convex set K is called a **cone** if

$$\boldsymbol{x} \in K \quad \Rightarrow \quad \lambda \boldsymbol{x} \in K \qquad \text{for all } \lambda \geq 0;$$

the **dual cone** to K is

$$K^{\dagger} = \{\boldsymbol{y} \,:\, \boldsymbol{y}^{\intercal}\boldsymbol{x} \geq 0 \text{ for all } \boldsymbol{x} \text{ in } K\}.$$

– and we have $(K^{\dagger})^{\dagger} = \operatorname{cl} K$

A convex set X induces a **support function**

$$\sigma_X(\boldsymbol{\lambda}) = \sup\{\boldsymbol{\lambda}^{\intercal}\boldsymbol{x} \,:\, \boldsymbol{x} \text{ in } X\}.$$

The **relative entropy function** continuously extends

$$D(\boldsymbol{u}, \boldsymbol{v}) = \sum_{i=1}^{m} u_i \log(u_i/v_i) \quad \text{to} \quad \mathbb{R}^m_+ \times \mathbb{R}^m_+.$$

Important: if you evaluate $D(\cdot,\cdot)$ outside $\mathbb{R}^m_+ \times \mathbb{R}^m_+$, you get $+\infty$.

A trick with convex duality

Caltech

Start with a **primal** problem

$$\mathrm{Val}(\boldsymbol{c}) = \inf_{\boldsymbol{x}}\{\boldsymbol{c}^\intercal \boldsymbol{x} \,:\, \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geq \boldsymbol{0}\}.$$

Obtain a **dual** problem

$$\mathrm{Val}(\boldsymbol{c}) = \sup_{\boldsymbol{\mu}}\{-\boldsymbol{b}^\intercal \boldsymbol{\mu} \,:\, \boldsymbol{A}^\intercal \boldsymbol{\mu} + \boldsymbol{c} \geq \boldsymbol{0}\}.$$

We will encounter constraints like

$$\mathrm{Val}(\boldsymbol{c}) + L \geq 0.$$

Write such a constraint as: *there exists* a $\boldsymbol{\mu}$ where

$$\boldsymbol{A}^\intercal \boldsymbol{\mu} + \boldsymbol{c} \geq \boldsymbol{0} \quad \text{and} \quad \boldsymbol{b}^\intercal \boldsymbol{\mu} \leq L.$$

Nonnegativity and optimization

Caltech

We'll work with sets $X \subset \mathbb{R}^n$. Speaking abstractly, for *any* $f : \mathbb{R}^n \to \mathbb{R}$

$$\begin{aligned} f_X^\star &= \inf\{f(\boldsymbol{x}) : \boldsymbol{x} \text{ in } X\} \\ &= \sup\{\gamma : f - \gamma \text{ is nonnegative over } X\}. \end{aligned}$$

Make this more concrete. For signomials:

$$\mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X) \doteq \left\{ \boldsymbol{c} \,:\, \sum_{i=1}^m c_i \exp(\boldsymbol{a}_i \cdot \boldsymbol{x}) \geq 0 \,\forall\, \boldsymbol{x} \in X \right\}, \quad \boldsymbol{A} \doteq \begin{bmatrix} \text{—} & \boldsymbol{0} & \text{—} \\ \text{—} & \boldsymbol{a}_2 & \text{—} \\ & \vdots & \\ \text{—} & \boldsymbol{a}_m & \text{—} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

So for $f(\boldsymbol{x}) = \sum_{i=1}^m c_i \exp(\boldsymbol{a}_i \cdot \boldsymbol{x})$,

$$f_X^\star = \sup\left\{\gamma \,:\, \boldsymbol{c} - \gamma \boldsymbol{e}_1 \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)\right\}$$

– where $\boldsymbol{e}_1$ is the 1^{st} standard basis vector in $\mathbb{R}^m$.

Duality and moment relaxations

Caltech

Abbreviate $\exp(\boldsymbol{A}\boldsymbol{x}) \in \mathbb{R}^m$ elementwise, and express

$$\mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X) = \{\boldsymbol{c} \,:\, \boldsymbol{c}^\intercal \exp(\boldsymbol{A}\boldsymbol{x}) \geq 0 \,\forall\, \boldsymbol{x} \in X\}.$$

The definition of "dual cone" requires

$$\mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)^\dagger = \{\boldsymbol{v} \,:\, \boldsymbol{c}^\intercal \boldsymbol{v} \geq 0 \,\forall\, \boldsymbol{c} \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)\}.$$

So we end up getting $\mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)^\dagger = \mathrm{co}\,\{\exp(\boldsymbol{A}\boldsymbol{x}) \,:\, \boldsymbol{x} \in X\}$ – a "**moment cone**."

$$\mathrm{conv}\,\{\exp(\boldsymbol{A}\boldsymbol{x}) \,:\, \boldsymbol{x} \in X\} = \left\{ \mathbb{E}_{\boldsymbol{x}}\left[\exp(\boldsymbol{A}\boldsymbol{x})\right] \,:\, \underbrace{\boldsymbol{x} \sim F, \;\; \mathrm{supp}\, F \subset X}_{\textit{conditional probability}} \right\}$$

Get *moment relaxations* from conic duality

$$\underbrace{\sup_{\gamma} \{\gamma \,:\, \boldsymbol{c} - \gamma \boldsymbol{e}_1 \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)\}}_{f_X^\star} = \inf \left\{ \boldsymbol{c}^\intercal \boldsymbol{v} \,:\, \begin{array}{c} \boldsymbol{v} \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)^\dagger \\ \text{satisfies } \boldsymbol{v} \cdot \boldsymbol{e}_1 = 1 \end{array} \right\}.$$

Motivation.

Mathematical Preliminaries.

Sums-of-AM/GM-Exponentials.

Sparsity preservation.

A hierarchy.

Extreme rays.

Conclusion.

X-AGE functions

Caltech

Definition. An X-**AGE** function is an X-nonnegative signomial, which has at most one negative coefficient. Generalizes $X = \mathbb{R}^n$ from [44]; see [45].

Consider $f(\boldsymbol{x}) = \sum_{i=1}^{m} c_i \exp(\boldsymbol{a}_i^\intercal \boldsymbol{x})$. If $\boldsymbol{c} \in \mathbb{R}^m$ has $\boldsymbol{c}_{\setminus k} \doteq (c_i)_{i \in [m]\setminus k} \geq \mathbf{0}$, then

$$f(\boldsymbol{x}) \geq 0 \quad \Leftrightarrow \quad \sum_{i=1}^{m} c_i \exp([\boldsymbol{a}_i - \boldsymbol{a}_k] \cdot \boldsymbol{x}) \geq 0 \quad \Leftrightarrow \quad \underbrace{\sum_{i \neq k} c_i \exp([\boldsymbol{a}_i - \boldsymbol{a}_k] \cdot \boldsymbol{x})}_{convex!} + c_k \geq 0.$$

Theorem (M., Chandrasekaran, & Wierman (2019))

If X is a convex set, then the conditions

$$\boldsymbol{c}_{\setminus k} \geq \mathbf{0} \quad \textit{and} \quad \boldsymbol{c} \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)$$

are equivalent to the existence of some $\boldsymbol{\nu} \in \mathbb{R}^m$ satisfying

$$\mathbf{1}^\intercal \boldsymbol{\nu} = 0 \quad \textit{and} \quad \sigma_X\left(-\boldsymbol{A}^\intercal \boldsymbol{\nu}\right) + D(\boldsymbol{\nu}_{\setminus k}, e\boldsymbol{c}_{\setminus k}) \leq c_k.$$

X-SAGE certificates & lower bounds

Caltech

A signomial is X-**SAGE** if it can be written as a sum of appropriate X-AGE functions.

The cone of coefficients

$$\mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, X) \doteq \Big\{ \boldsymbol{c} : \{(\boldsymbol{\nu}^{(k)}, \boldsymbol{c}^{(k)})\}_{k=1}^{m} \text{ satisfy } \boldsymbol{c} = \textstyle\sum_{k=1}^{m} \boldsymbol{c}^{(k)},\ \mathbf{1}^\intercal \boldsymbol{\nu}^{(k)} = 0, \\ \text{and } \sigma_X\left(-\boldsymbol{A}\boldsymbol{\nu}^{(k)}\right) + D\left(\boldsymbol{\nu}^{(k)}_{\setminus k}, \boldsymbol{c}^{(k)}_{\setminus k}\right) \leq c^{(k)}_k \ \forall\, k \in [m] \Big\}$$

is contained within $\mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)$.

Consider $f(\boldsymbol{x}) = \sum_{i=1}^{m} c_i \exp(\boldsymbol{a}_i \cdot \boldsymbol{x})$ with $\boldsymbol{a}_1 = \mathbf{0}$:

$$\begin{aligned} f_X^\star &= \sup\{\gamma \,:\, \boldsymbol{c} - \gamma \boldsymbol{e}_1 \text{ in } \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)\} \\ &\geq \sup\{\gamma \,:\, \boldsymbol{c} - \gamma \boldsymbol{e}_1 \text{ in } \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, X)\} =: f_X^{\mathsf{SAGE}}. \end{aligned}$$

MOSEK + `sageopt` = off-the-shelf software for computing f_X^{SAGE}.

Conditional moment relaxations via SAGE

Caltech

Consider $f(\boldsymbol{x}) = \sum_{i=1}^{m} c_i \exp(\boldsymbol{a}_i \cdot \boldsymbol{x})$ with $\boldsymbol{a}_1 = \boldsymbol{0}$. Applying conic duality ...

$$\sup\{\, \gamma \;:\; \boldsymbol{c} - \gamma \boldsymbol{e}_1 \text{ in } \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, X)\} = f_X^{\mathsf{SAGE}} = \inf\left\{ \boldsymbol{c}^\intercal \boldsymbol{v} \;:\; \begin{array}{c} \boldsymbol{v} \text{ in } \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, X)^\dagger \\ \text{satisfies } \boldsymbol{v} \cdot \boldsymbol{e}_1 = 1 \end{array} \right\}$$

Conic duality reverses inclusions

$$\mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)^\dagger \subset \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, X)^\dagger.$$

The dual X-SAGE cone is

$$\begin{aligned} \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, X)^\dagger = \mathrm{cl}\{\boldsymbol{v} : \; & \text{some } \boldsymbol{z}_1, \ldots, \boldsymbol{z}_m \text{ in } \mathbb{R}^n \text{ satisfy} \\ & v_k \log(\boldsymbol{v}/v_k) \geq [\boldsymbol{A} - \boldsymbol{1}\boldsymbol{a}_k]\boldsymbol{z}_k \\ & \text{and } \boldsymbol{z}_k / v_k \in X \text{ for all } k \text{ in } [m]\}. \end{aligned}$$

The dual helps with solution recovery. Useful even when $f_X^{\mathsf{SAGE}} < f_X^\star$!

An example in $\mathbb{R}^3$

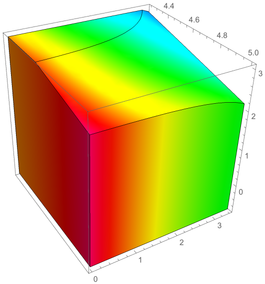
Caltech

Minimize

$$f(\boldsymbol{x}) = \exp(x_1 - x_2)/2 - \exp x_1 - 5\exp(-x_2)$$

over

$$X = \left\{\boldsymbol{x} : (70,\ 1,\ 0.5) \leq \exp \boldsymbol{x} \leq (150,\ 30,\ 21)\right.$$
$$\left.\frac{\exp(x_2 - x_3)}{100} + \frac{\exp x_2}{100} + \frac{\exp(x_1 + x_3)}{2000} \leq 1\right\}.$$



Compute $f_X^{\mathsf{SAGE}} = -147.85713 \leq f_X^\star$, and recover feasible

$$\tilde{\boldsymbol{x}} = (5.01063529,\ 3.40119660, -0.48450710)$$

satisfying $f(\tilde{\boldsymbol{x}}) = -147.66666$.

Sparsity and SAGE signomials

Caltech

Theorem (M., Chandrasekaran, & Wierman)

Fix a vector $\boldsymbol{c} \in \mathbb{R}^m$ with nonempty $N = \{i : c_i < 0\}$.

If $\boldsymbol{c} \in \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, X)$, then there exist X-AGE vectors $\{\boldsymbol{c}^{(i)}\}_{i \in N}$ where $c_i^{(i)} = c_i$ and

$$\boldsymbol{c} = \sum_{i \in N} \boldsymbol{c}^{(i)}.$$

This is true even if X is not convex.

Proven formally for $X = \mathbb{R}^n$ in [46].

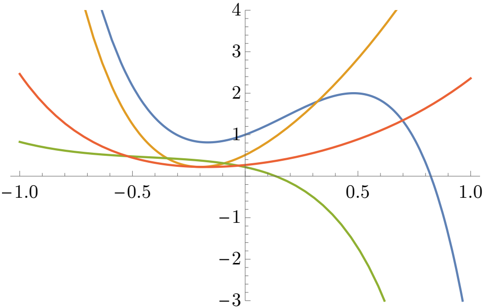
Let $\mathsf{K} \supset \mathbb{R}^m_+$ induce $\mathsf{C}_i \doteq \{\boldsymbol{c} \in \mathsf{K} : \boldsymbol{c}_{\setminus i} \geq \mathbf{0}\}$.

1. Show that for $j \notin N$, can eliminate $\boldsymbol{c}^{(j)} \in \mathsf{C}_j$ from an existing decomposition.
2. Show that conic combinations of $\{\boldsymbol{c}^{(i)} \in \mathsf{C}_i\}_{i \in N}$ can reduce to $c_i^{(i)} = c_i$.

Sparsity preservation: a univariate example

Caltech

$$f(x) = e^{-3x} + e^{-2x} + 4e^{x} + e^{2x} - 4e^{-x} - 1 - e^{3x} \text{ over } x \leq 0$$



$$f_1(x) = 0.88 \cdot e^{-3x} + 0.82 \cdot e^{-2x} + 2.69 \cdot e^{x} + 0.12 \cdot e^{2x} - 4 \cdot e^{-x}$$
$$f_2(x) = 0.10 \cdot e^{-3x} + 0.15 \cdot e^{-2x} + 0.90 \cdot e^{x} + 0.12 \cdot e^{2x} - 1$$
$$f_3(x) = 0.02 \cdot e^{-3x} + 0.03 \cdot e^{-2x} + 0.41 \cdot e^{x} + 0.76 \cdot e^{2x} - e^{3x}$$

Applying Sums-of-Squares (SOS) to an AGE function

Consider the following function on $\mathbb{R}^2$

$$f(x, y) = 1 - 2e^{(2x+2y)} + \frac{1}{2}\left(e^{8x} + e^{8y}\right).$$

Use sageopt, round solution, certify f is $\mathbb{R}^2$-AGE with $\boldsymbol{\nu}^\star = (1, -2, 1/2, 1/2)$.

We can express f as a sum-of-squares, but this requires new terms

$$\begin{aligned} f(x, y) &= \left(1 - 2e^{(2x+2y)}\right)^2 + \frac{1}{2}\left(e^{4x} - e^{4y}\right)^2 \\ &= \left(1 - 2e^{(2x+2y)} + e^{(4x+4y)}\right) + \frac{1}{2}\left(e^{8x} + e^{8y} - 2e^{(4x+4y)}\right). \end{aligned}$$

SOS is the predominant way to certify polynomial nonnegativity.

SAGE can certify polynomial nonnegativity [46] with $X \subsetneq \mathbb{R}^n$ [45].

Remark: In the special case $X = \mathbb{R}^n$ with integer exponents, the sparsity result can also be deduced from Jie Wang's work on Sums-of-Nonnegative-Circuit polynomials [47, 48].

Motivation.

Mathematical Preliminaries.

Sums-of-AM/GM-Exponentials.

Sparsity preservation.

A hierarchy.

Extreme rays.

Conclusion.

A hierarchy of stronger convex relaxations.

Caltech

The earlier example with $X \subset \mathbb{R}^3$ –

$$f(\boldsymbol{x}) = \exp(x_1 - x_2)/2 - \exp x_1 - 5\exp(-x_2).$$

We found bounds

$$f_X^{\mathsf{SAGE}} = -147.85713 \leq f_X^\star \quad \text{and} \quad f_X^\star \leq f(\tilde{\boldsymbol{x}}) = -147.66666.$$

The *modulation trick* lets us construct a sequence of bounds

$$f_X^{(\ell)} \doteq \sup\{\gamma \,:\, \left(\textstyle\sum_{i=1}^m \exp(\boldsymbol{a}_i \cdot \boldsymbol{x})\right)^\ell (f(\boldsymbol{x}) - \gamma) \text{ is } X\text{-SAGE}\}.$$

Using MOSEK + sageopt with this particular example,

ℓ	SAGE bound	solve time (s)
0	-147.85713	0.01
1	-147.67225	0.02
2	-147.66680	0.08
3	-147.66666	0.26

Convergence results

When introducing $\mathbb{R}^n$-SAGE, Chandrasekaran and Shah proved two results for *hierarchies*.

No known convergence conditions for $f_X^{(\ell)}$, prior to March 2020.

Theorem (Wang, Jaini, Yu, Poupart [49])

Let $\mathbf{A} \in \mathbb{Q}^{m \times n}$ be a rank n matrix with $\mathbf{a}_1 = \mathbf{0}$, and consider $f(\mathbf{x}) = \sum_{i=1}^m c_i \exp(\mathbf{a}_i \cdot \mathbf{x})$. If X is a compact convex set, then

$$\lim_{\ell \to \infty} f_X^{(\ell)} = f_X^\star.$$

Assumes nothing about the representation of X.

Compare to the canonical (non X-SAGE) approach, which uses a Lagrangian relaxation:

$$\inf_{\mathbf{x}}\{f(\mathbf{x}) \,:\, g(\mathbf{x}) \geq \mathbf{0}\} \geq \sup_{\gamma, \boldsymbol{\lambda}}\{\gamma \,:\, f - \gamma - \textstyle\sum_i \lambda_i \cdot g_i \in \Lambda,\ \lambda_i \in \Lambda'\}$$

Λ, Λ' are sets of nonnegative functions.

Motivation.

Mathematical Preliminaries.

Sums-of-AM/GM-Exponentials.

Sparsity preservation.

A hierarchy.

Extreme rays.

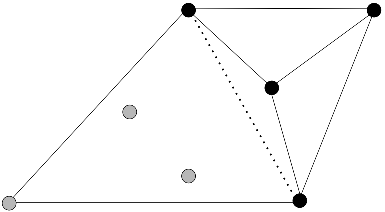
Conclusion.

Affine matroid-theoretic circuits

Caltech

A **circuit** is a minimal affinely dependent $\{\boldsymbol{a}_i\}_{i \in I} \subset \mathbb{R}^n$.

A circuit is **simplicial** if $\operatorname{conv}\{\boldsymbol{a}_i\}_{i \in I}$ has $|I| - 1$ extreme points.



As a matter of notation, let $\mathcal{L}^k = \{\boldsymbol{\lambda} \,:\, \boldsymbol{\lambda}^\intercal \mathbf{1} = 0,\ \lambda_k = -1,\ \boldsymbol{\lambda}_{\setminus k} \geq \mathbf{0}\}$.

Simplicial circuits obtained from $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ are 1-to-1 with certain $\boldsymbol{\lambda} \in \mathbb{R}^m$

$$\boldsymbol{\lambda} \in \mathcal{L}^k \text{ for some } k \in [m], \quad \boldsymbol{A}^\intercal \boldsymbol{\lambda} = \mathbf{0} \quad \text{and} \quad \{\boldsymbol{a}_i \,:\, \lambda_i > 0\} \text{ is affinely independent.}$$

Circuits and $\mathbb{R}^n$-SAGE

Caltech

M., Chandrasekaran, and Wierman [46] determined extreme rays of

$$\mathsf{C}_{\mathsf{AGE}}(\boldsymbol{A}, i) = \{\boldsymbol{c} \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, \mathbb{R}^n) \,:\, \boldsymbol{c}_{\setminus i} \geq \mathbf{0}\}.$$

The ordinary SAGE cone is a Minkowski sum

$$\mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}) = \sum_{i=1}^{m} \mathsf{C}_{\mathsf{AGE}}(\boldsymbol{A}, i).$$

Katthän, Naumann, and Theobald [50] completely determined $\operatorname{ext} \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A})$.

Forsgård and de Wolff [51] studied $\partial \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A})$ in *detail*;defined

$$\mathrm{Rez}(\boldsymbol{A}) = \mathrm{co}\{\boldsymbol{\lambda} \in \mathbb{R}^m \,:\, \boldsymbol{\lambda} \text{ is a simplicial circuit w.r.t. } \boldsymbol{A}\}.$$

Combine [50, 51] to clearly link $\operatorname{ext} \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A})$ and $\operatorname{ext} \mathrm{Rez}(\boldsymbol{A})$.

Circuits and X-SAGE

Caltech

The following is ongoing, joint work with Helen Naumann and Thorsten Theobald.

Definition. A simplicial X-**circuit** induced by $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a vector $\boldsymbol{\lambda}^\star \in \mathbb{R}^m$ where

1. $\boldsymbol{\lambda}^\star \in \mathcal{L}_k$ for some $k \in [m]$,
2. $\sigma_X\left(-\boldsymbol{A}^\intercal \boldsymbol{\lambda}^\star\right) < +\infty$, and
3. if $\boldsymbol{\lambda} \mapsto \sigma_X\left(-\boldsymbol{A}^\intercal \boldsymbol{\lambda}\right)$ is linear on $[\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2] \subset \mathcal{L}^k$, then $\boldsymbol{\lambda}^\star \notin \operatorname{relint}[\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2]$.

A clean generalization from $X = \mathbb{R}^n$.

Provides the basis for a "Reznick cone" with conditional SAGE certificates.

Particularly informative when X is a polyhedron.

E.g., if X is a polyhedron, then $\mathrm{C}_{\mathrm{SAGE}}(\boldsymbol{A}, X)$ is power-cone representable.

Motivation.

Mathematical Preliminaries.

Sums-of-AM/GM-Exponentials.

Sparsity preservation.

A hierarchy.

Extreme rays.

Conclusion.

Some open problems

Caltech

1. When do we have $\mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, X) = \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, X)$?

   For $X = \mathbb{R}^n$ see [46, 51], and also [46, 47] for polynomials.

2. If $f > 0$ on compact X, is there some $g > 0$ so $f \cdot g$ is X-SAGE?

   Resolved in the affirmative for $\boldsymbol{A} \in \mathbb{Q}^{m \times n}$ [49]! *Follow-up questions ...*

   If "$h =$standard multiplier," how to bound least ℓ where $h^\ell \cdot f$ is X-SAGE?

   Irrational $\boldsymbol{A}$? Perhaps leverage Hausdorff continuity.

3. Complexity of testing "$\boldsymbol{c} \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{\alpha}, X)$" with *two* $c_i < 0$?

Many possible algorithmic projects (ask me for details).

More open problems to follow once "X-circuit" paper is put on arXiv.

Acknowledgements

Caltech

Special thanks to my collaborators, on all things SAGE-related:

Venkat Chandrasekaran, Adam Wierman, Helen Naumann, Thorsten Theobald, Fang Xiao, and Berk Orztuk.

Thanks also to the many people who make working in this area exciting:

Sadik Iliman, Timo de Wolff, Marieke Dressler, Henning Seidler, Janin Heuer, Jie Wang, Jen Forsgård, and more!

NSF

Max-Planck-Institut für Mathematik in den Naturwissenschaften

References I

Caltech

[1] Yurii Nesterov and Arkadii Nemirovskii.
Interior-Point Polynomial Algorithms in Convex Programming.
Society for Industrial and Applied Mathematics, January 1994.

[2] Stephen P. Boyd, Seung-Jean Kim, Dinesh D. Patil, and Mark A. Horowitz.
Digital circuit optimization via geometric programming.
Operations Research, 53(6):899–932, December 2005.

[3] R. A. Jabr.
Application of geometric programming to transformer design.
IEEE Transactions on Magnetics, 41(11):4261–4269, 2005.

[4] D. Patil, S. Yun, S. . Kim, A. Cheung, M. Horowitz, and S. Boyd.
A new method for design of robust digital circuits.
In *Sixth international symposium on quality electronic design (isqed'05)*, pages 676–681, 2005.

[5] Ting-Yu Chen.
Structural optimization using single-term posynomial geometric programming.
Computers & Structures, 45(5-6):911–918, December 1992.

References II

Caltech

[6] Prabhat Hajela.
Geometric programming strategies in large-scale structural synthesis.
AIAA Journal, 24(7):1173–1178, July 1986.

[7] Yves Smeers and Daniel Tyteca.
A geometric programming model for the optimal design of wastewater treatment plants.
Operations Research, 32(2):314–342, April 1984.

[8] Warren Hoburg and Pieter Abbeel.
Geometric programming for aircraft design optimization.
AIAA Journal, 52(11):2414–2426, November 2014.

[9] Michael Burton and Warren Hoburg.
Solar and gas powered long-endurance unmanned aircraft sizing via geometric programming.
Journal of Aircraft, 55(1):212–225, January 2018.

[10] C. Nowzari, V. M. Preciado, and G. J. Pappas.
Optimal resource allocation for control of networked epidemic models.
IEEE Transactions on Control of Network Systems, 4(2):159–169, 2017.

References III

[11] V. M. Preciado, M. Zargham, C. Enyioha, A. Jadbabaie, and G. J. Pappas.
Optimal resource allocation for network protection against spreading processes.
IEEE Transactions on Control of Network Systems, 1(1):99–108, 2014.

[12] C. Nowzari, V. M. Preciado, and G. J. Pappas.
Analysis and control of epidemics: A survey of spreading processes on complex networks.
IEEE Control Systems Magazine, 36(1):26–46, 2016.

[13] M. Chiang, C. W. Tan, D. P. Palomar, D. O'neill, and D. Julian.
Power control by geometric programming.
IEEE Transactions on Wireless Communications, 6(7):2640–2651, 2007.

[14] J. Rajasekharan and V. Koivunen.
Optimal energy consumption model for smart grid households with energy storage.
IEEE Journal of Selected Topics in Signal Processing, 8(6):1154–1166, 2014.

References IV

Caltech

[15] Abhijit Davare, Qi Zhu, Marco Di Natale, Claudio Pinello, Sri Kanajan, and Alberto Sangiovanni-Vincentelli.
Period optimization for hard real-time distributed automotive systems.
In *Proceedings of the 44th Annual Design Automation Conference*, DAC '07, page 278–283, New York, NY, USA, 2007. Association for Computing Machinery.

[16] S. Misra, M. W. Fisher, S. Backhaus, R. Bent, M. Chertkov, and F. Pan.
Optimal compression in natural gas networks: A geometric programming approach.
IEEE Transactions on Control of Network Systems, 2(1):47–56, 2015.

[17] Y. He, W. Zhu, and L. Guan.
Optimal resource allocation for pervasive health monitoring systems with body sensor networks.
IEEE Transactions on Mobile Computing, 10(11):1558–1575, 2011.

[18] Alberto Marin-Sanguino, Eberhard O Voit, Carlos Gonzalez-Alcon, and Nestor V Torres.
Optimization of biotechnological systems through geometric programming.
Theoretical Biology and Medical Modelling, 4(1), September 2007.

References V

Caltech

[19] J.R. Rajasekera and M. Yamada.
Estimating the firm value distribution function by entropy optimization and geometric programming.
Annals of Operations Research, 105(1/4):61–75, 2001.

[20] Farshid Samadi, Abolfazl Mirzazadeh, and Mir Mohsen Pedram.
Fuzzy pricing, marketing and service planning in a fuzzy inventory model: A geometric programming approach.
Applied Mathematical Modelling, 37(10-11):6683–6694, June 2013.

[21] Seyed Ahmad Yazdian, Kamran Shahanaghi, and Ahmad Makui.
Joint optimisation of price, warranty and recovery planning in remanufacturing of used products under linear and non-linear demand, return and cost functions.
International Journal of Systems Science, 47(5):1155–1175, May 2014.

[22] M. MAZUMDAR and T. R. JEFFERSON.
Maximum likelihood estimates for multinomial probabilities via geometric programming.
Biometrika, 70(1):257–261, 04 1983.

References VI

Caltech

[23] Paul D. Feigin and Ury Passy.
The geometric programming dual to the extinction probability problem in simple branching processes.
The Annals of Probability, 9(3):498–503, 1981.

[24] Martin A. York, Warren W. Hoburg, and Mark Drela.
Turbofan engine sizing and tradeoff analysis via signomial programming.
Journal of Aircraft, 55(3):988–1003, May 2018.

[25] Martin A. York, Berk Öztürk, Edward Burnell, and Warren W. Hoburg.
Efficient aircraft multidisciplinary design optimization and sensitivity analysis via signomial programming.
AIAA Journal, 56(11):4546–4561, November 2018.

[26] Berk Ozturk and Ali Saab.
Optimal aircraft design deicions under uncertainty via robust signomial programming.
In *AIAA Aviation 2019 Forum*. American Institute of Aeronautics and Astronautics, June 2019.

References VII

Caltech

[27] Philippe G. Kirschen and Warren W. Hoburg.
The power of log transformation: A comparison of geometric and signomial programming with general nonlinear programming techniques for aircraft design optimization.
In *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. American Institute of Aeronautics and Astronautics, January 2018.

[28] Philippe G. Kirschen, Martin A. York, Berk Ozturk, and Warren W. Hoburg.
Application of signomial programming to aircraft design.
Journal of Aircraft, 55(3):965–987, May 2018.

[29] M. Avriel and J. D. Barrett.
Optimal design of pitched laminated wood beams.
In *Advances in Geometric Programming*, pages 407–419. Springer US, 1980.

[30] Hojjat Adeli and Osama Kamal.
Efficient optimization of space trusses.
Computers & Structures, 24(3):501 – 511, 1986.

References VIII

Caltech

[31] Balaur S. Dhillon and Chen-Hsing Kuo.
Optimum design of composite hybrid plate girders.
Journal of Structural Engineering, 117(7):2088–2098, 1991.

[32] Yun Kang Sui and Xi Cheng Wang.
Second-order method of generalized geometric programming for spatial frame optimization.
Computer Methods in Applied Mechanics and Engineering, 141(1-2):117–123, February 1997.

[33] R.A. Jabr.
Inductor design using signomial programming.
COMPEL - The international journal for computation and mathematics in electrical and electronic engineering, 26(2):461–475, 2007.

[34] Mung Chiang.
Nonconvex Optimization for Communication Networks, pages 137–196.
Springer US, Boston, MA, 2009.

References IX

Caltech

[35] Han Zhao, Pascal Poupart, and Geoffrey J Gordon.
A unified approach for learning the parameters of sum-product networks.
In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 433–441. Curran Associates, Inc., 2016.

[36] Bruce Reznick.
Extremal PSD forms with few terms.
Duke Mathematical Journal, 45(2):363–374, jun 1978.

[37] Bruce Reznick.
Forms derived from the arithmetic-geometric inequality.
Mathematische Annalen, 283(3):431–464, 1989.

[38] Philippe P. Pébay, J. Maurice Rojas, and David C. Thompson.
Optimization and NP_r-completeness of certain fewnomials.
In *Proceedings of the 2009 conference on Symbolic numeric computation*. ACM Press, 2009.

References X

Caltech

[39] Mehdi Ghasemi and Murray Marshall.
Lower bounds for polynomials using geometric programming.
SIAM Journal on Optimization, 22(2):460–473, jan 2012.

[40] M. Ghasemi, J. B. Lasserre, and M. Marshall.
Lower bounds on the global minimum of a polynomial.
Computational Optimization and Applications, 57(2):387–402, sep 2013.

[41] Casian Pantea, Heinz Koeppl, and Gheorghe Craciun.
Global injectivity and multiple equilibria in uni- and bi-molecular reaction networks.
Discrete and Continuous Dynamical Systems - Series B, 17(6):2153–2170, May 2012.

[42] Elias August, Gheorghe Craciun, and Heinz Koeppl.
Finding invariant sets for biological systems using monomial domination.
In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, December 2012.

[43] Sadik Iliman and Timo de Wolff.
Amoebas, nonnegative polynomials and sums of squares supported on circuits.
Research in the Mathematical Sciences, 3(1):9, Mar 2016.

References XI

Caltech

[44] Venkat Chandrasekaran and Parikshit Shah.
Relative entropy relaxations for signomial optimization.
SIAM Journal on Optimization, 26(2):1147 – 1173, 2016.

[45] Riley Murray, Venkat Chandrasekaran, and Adam Wierman.
Signomial and polynomial optimization via relative entropy and partial dualization, 2019.

[46] Riley Murray, Venkat Chandrasekaran, and Adam Wierman.
Newton polytopes and relative entropy optimization, 2018.

[47] Jie Wang.
Nonnegative polynomials and circuit polynomials, 2018.

[48] Jie Wang.
On supports of sums of nonnegative circuit polynomials, 2018.

[49] Allen Houze Wang, Priyank Jaini, Yaoliang Yu, and Pascal Poupart.
Complete hierarchy of relaxation for constrained signomial positivity, 2020.

References XII

Caltech

[50] L. Katthän, H. Naumann, and T. Theobald.
A unified framework of SAGE and SONC polynomials and its duality theory, 2019.
Preprint, arXiv:1903.08966.

[51] Jens Forsgård and Timo de Wolff.
The algebraic boundary of the sonc cone, 2019.

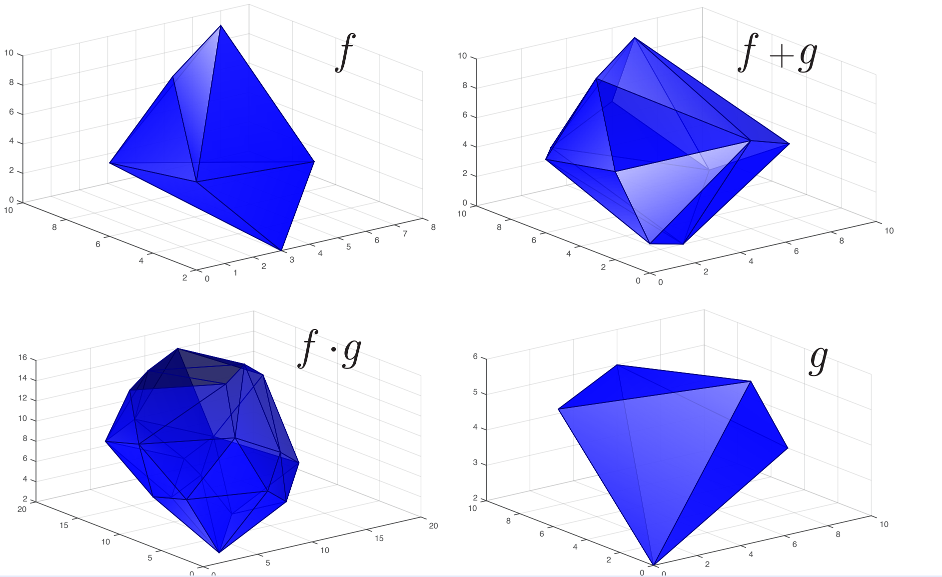
Appendices

Caltech

- Exactness analysis.
- Software.
- Optimization with nonconvex constraints.
- Log-log convex ("geometrically convex") functions.

$\mathbb{R}^n$-SAGE Exactness

Caltech

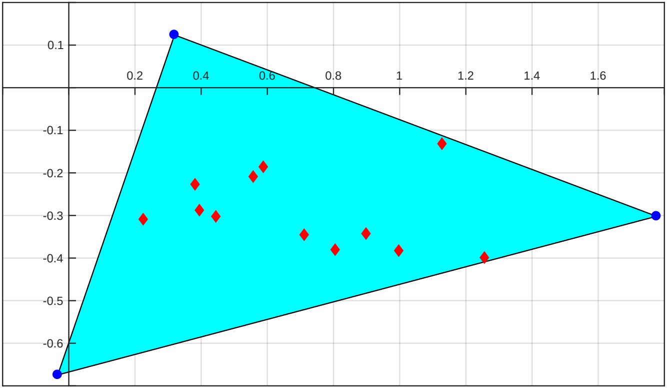


Simplicial sign patterns

Caltech

Theorem (7)

If $\mathrm{conv}(\boldsymbol{A})$ *is simplicial, and* $c_i \leq 0$ *for all nonextremal* $\boldsymbol{a}_i$*,*
then $\boldsymbol{c} \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, \mathbb{R}^n)$ *if and only if* $\boldsymbol{c} \in \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, \mathbb{R}^n)$.

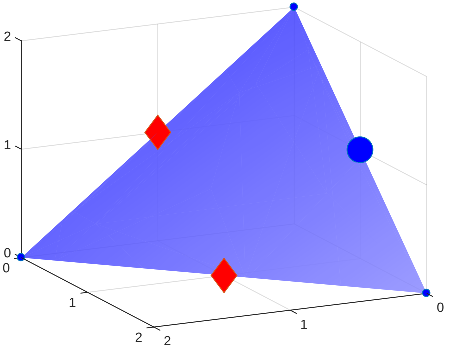


Simplicial sign patterns

Caltech

Theorem (7)

If $\mathrm{conv}(\boldsymbol{A})$ *is simplicial, and* $c_i \leq 0$ *for all nonextremal* $\boldsymbol{a}_i$*,*
then $\boldsymbol{c} \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, \mathbb{R}^n)$ *if and only if* $\boldsymbol{c} \in \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, \mathbb{R}^n)$.

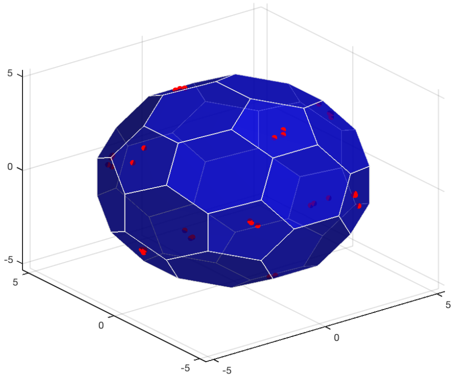


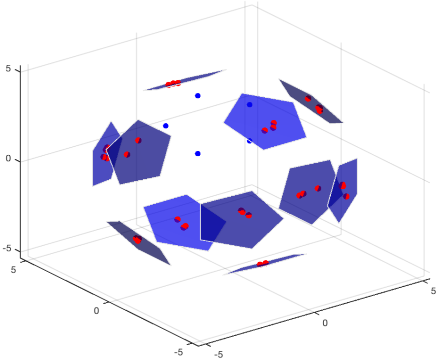
$$f(\boldsymbol{x}) = (e^{x_1} - e^{x_2} - e^{x_3})^2$$

Partitioning a Newton polytope

Caltech

We say that $\boldsymbol{A}$ can be **partitioned into ℓ faces** if we can permute its rows so that $\boldsymbol{A} = [\boldsymbol{A}^{(1)}; \ldots; \boldsymbol{A}^{(\ell)}]$ where $\{\mathrm{conv}\, \boldsymbol{A}^{(i)}\}_{i=1}^{\ell}$ are mutually disjoint faces of $\mathrm{conv}(\boldsymbol{A})$.





Partitioning a Newton polytope

Caltech

Theorem (8)

If $\{\boldsymbol{A}^{(i)}\}_{i=1}^{\ell}$ *are matrices partitioning* $\boldsymbol{A} = [\boldsymbol{A}^{(1)}; \ldots; \boldsymbol{A}^{(\ell)}]$*, then*

$$\mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, \mathbb{R}^n) = \oplus_{i=1}^{\ell} \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}^{(i)}, \mathbb{R}^n)$$

–and the same is true of $\mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, \mathbb{R}^n)$*.*

Sanity checks :

All matrices $\boldsymbol{A}$ admit a trivial partition with $\ell = 1$.

If all $\boldsymbol{a}_i$ are extremal, then $\mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, \mathbb{R}^n) = \mathbb{R}^m_+$.

An $\mathbb{R}^n$-SAGE exactness theorem

Caltech

Theorem (9)

Suppose $\boldsymbol{A}$ can be partitioned into faces where

1. *each simplicial face has ≤ 2 nonextremal exponents, and*
2. *all other faces contain at most one nonextremal exponent.*

Then $\mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, \mathbb{R}^n) = \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, \mathbb{R}^n)$.

Violate the first hypothesis? Consider

$$f(\boldsymbol{x}) = (e^{x_1} - e^{x_2} - e^{x_3})^2 \qquad \text{not SAGE, per C\&S'16.}$$

Violate the second hypothesis? Consider

$$\boldsymbol{A}^\intercal = [\boldsymbol{e}_1,\ \boldsymbol{e}_2,\ 2\boldsymbol{e}_1,\ 2\boldsymbol{e}_2,\ 2(\boldsymbol{e}_1 + \boldsymbol{e}_2),\ \boldsymbol{0}],$$

for which $(-4,\ -2,\ 3,\ 2,\ 1,\ 1.8) \in \mathsf{C}_{\mathsf{NNS}}(\boldsymbol{A}, \mathbb{R}^2) \setminus \mathsf{C}_{\mathsf{SAGE}}(\boldsymbol{A}, \mathbb{R}^2)$.

Optimization with nonconvex constraints

Caltech

Q: What should we do when some constraints are nonconvex?

A: Combine X-SAGE certificates with Lagrangian relaxations.

Concretely, suppose we want to minimize f over

$$\Omega \doteq X \cap \{\boldsymbol{x} \,:\, g(\boldsymbol{x}) \leq \mathbf{0}\}$$

where X is convex, but $g_1, \ldots, g_k$ are nonconvex signomials.

Then, if $\lambda_1, \ldots, \lambda_k$ are nonnegative dual variables, we have

$$\inf_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}) \geq \sup \left\{ \gamma \,:\, f + \sum_{i=1}^{k} \lambda_i g_i - \gamma \text{ is } X\text{-SAGE} \right\}.$$

The `SimPleAC` aircraft design problem

Caltech

From Warren Hoburg's PhD thesis.

Problem statistics:

- 140 variables.
- 89 inequality constraints (1 nonconvex).
- 67 equality constraints (15 nonconvex).

Performance of the most basic SAGE relaxation:

- bound "cost ≥ 2957" (roughly match a known solution).
- MOSEK solves in two seconds, on a six year old laptop.
- solution recovery fails (numerical issues).

The sageopt python package

```
import sageopt as so

y = so.standard_sig_monomials(3)
f = 0.5*y[0]/y[1] - y[0] - 5/y[1]
ineqs = [100 -  y[1]/y[2] - y[1] - 0.05*y[0]*y[2],
         y[0] - 70, y[1] - 1, y[2] - 0.5,
         150 - y[0], 30 - y[1], 21 - y[2]]
X = so.infer_domain(f, gts, [])

prob = so.sig_relaxation(f, X, form='dual')
prob.solve()
solutions = so.sig_solrec(prob)
```

The sageopt python package

Caltech

```
...  # define f, X as before
from sageopt import coniclifts as cl

modulator = so.Signomial(f.alpha, np.ones(f.m)) ** 3
gamma = cl.Variable()
h = modulator * (f - gamma)
con = cl.PrimalSageCone(h.c, h.alpha, X, 'con_name')
prob = cl.Problem(cl.MAX, gamma, [con])

prob.solve()
age_vecs = [v.value for v in con.age_vectors.values()]
age_sigs = [so.Signomial(h.alpha, v) for v in age_vecs]
h_numeric = so.Signomial(h.alpha, h.c.value)
```

Log-log convexity: examples

Caltech

With domains $D = \mathbb{R}^n_{++}$:

$$g(\boldsymbol{x}) = \max\{x_1, \ldots, x_n\}$$

$$g(\boldsymbol{x}) = x_1^{a_1} \cdots x_n^{a_n}$$

$$g(x) = \left(\int_x^\infty e^{-t^2} \mathrm{d}t\right)^{-1}$$

With more restricted domains:

$$x \mapsto (-x \log x)^{-1} \qquad D = (0, 1)$$

$$\boldsymbol{X} \mapsto (\boldsymbol{I} - \boldsymbol{X})^{-1}$$

$$D = \{\boldsymbol{X} \in \mathbb{R}^{n \times n}_{++} \,:\, \rho(\boldsymbol{X}) < 1\}$$

$$x \mapsto (\log x)^{-1} \qquad D = (1, \infty)$$

Some tractable constraints for X-SAGE polynomials:

$$\|\boldsymbol{x}\|_p \leq a \qquad x_j^2 = a \qquad a \leq \mathbb{P}\{\mathcal{N}(0, \sigma) \geq |x|\}$$

where $a > 0$.