POp Reading group: Signomial Optimization via Relative Entropy Cone

Based on a paper by Venkat Chandrasekaran and Parikshit Shah

Presented by: Andries Steenkamp

Outline for Section 1

Definitions

Signomials

Definition

A function $g : \mathbb{R}^n \to \mathbb{R}$ is called a *signomial* if it is of the form

$$g(x) := \sum_{j=1}^{\ell} c_j \exp(\alpha^{(j)^T} x)$$

where $c_j \in \mathbb{R}$, $\alpha^{(j)} \in \mathbb{R}^n$ for all $j \in [\ell]$.

Definitions

Signomial Optimization

Definition

A *signomial program* is of the form

$$\begin{aligned} \inf \quad & g(x) \\ \text{s.t.} \quad & g_i(x) \geq 0 \; \forall \, i \in [m] \end{aligned}$$

where $g, g_1, \ldots, g_m$ are signomials.

Definitions

a Special case of signomials

Definition

A signomial $g(x)$ is called an *AM-GM-Exponential* if the following holds

- $g(x) \geq 0$ (globally nonnegative)
- All coefficients , c_j's, with the possible exception of one are positive.

$$g(x) := b \exp(\alpha^T x) + \sum_{j=1}^{\ell} c_j \exp(\alpha^{(j)^T} x)$$

where $b \in \mathbb{R}$, $c_j \in \mathbb{R}_{++}$, $\alpha, \alpha^{(j)} \in \mathbb{R}^n$ for all $j \in [\ell]$

Definitions

AM-GM and Relative Entropy

Definition

Let $w_1, \ldots, w_n > 0$, $\sum_{i=1}^n w_j = w$ and $x_1, \ldots, x_n \geq 0$ then the (weighted) *Arithmetic-Geometric mean inequality* is given by

$$\frac{1}{w} \sum_{i=1}^{n} w_i x_i \geq \sqrt[w]{\prod_{i=1}^{n} x_i^{w_i}}.$$

The *Relative Entropy* function is defined to be

$$D : \mathbb{R}^n \times \mathbb{R}^n \ni (v, c) \mapsto \sum_{i=1}^{n} v_i \log(\frac{v_i}{c_i}) \in \mathbb{R} \cup \{\infty, -\infty\}$$

Result

AM-GM-Exponentials and relative entropy

Lemma

Let $b \in \mathbb{R}$, $c_j \in \mathbb{R}_{++}$, $\alpha, \alpha^{(j)} \in \mathbb{R}^n$ for all $j \in [\ell]$ then

$$g(x) := b \exp(\alpha^T x) + \sum_{j=1}^{\ell} c_j \exp(\alpha^{(j)^T} x) \geq 0 \tag{1}$$

$$\Longleftrightarrow$$

$$\exists v \in \mathbb{R}^{\ell}_{+} \text{ s.t. } D(v, \exp(1)c) - b \leq 0 \tag{2.1}$$

$$[\alpha^{(1)}, \ldots, \alpha^{(\ell)}]v = \sum_{j=1}^{\ell} v_j \alpha^{(j)} = (\mathbb{1}^T v)\alpha \tag{2.2}$$

Proof

(2) $\implies$ *(1)*

Suppose you are give a v that satisfies (2.1) and (2.2).

$$\frac{1}{1}\sum_{j=1}^{\ell}\frac{v_j/\mathbb{1}^T v}{v_j/\mathbb{1}^T v}c_j\exp(\alpha^{(j)^T}x) \geq \left(\prod_{j=1}^{\ell}\left(\frac{c_j\exp(\alpha^{(j)^T}x)}{v_j/\mathbb{1}^T v}\right)^{v_j/\mathbb{1}^T v}\right)^1$$

$\because$ weighted AM-GM inequality

$$= \prod_{j=1}^{\ell}\left(\frac{c_j\exp(\frac{v_j\alpha^{(j)^T}x}{\mathbb{1}^T v})}{v_j/\mathbb{1}^T v}\right) = \prod_{j=1}^{\ell}\left(\frac{c_j}{v_j/\mathbb{1}^T v}\right)^{v_j/\mathbb{1}^T v}\exp(\alpha^T x)$$

$\because \sum_{j=1}^{\ell} v_j\alpha^{(j)} = (\mathbb{1}^T v)\alpha$ by (2.2)

Proof

(2) $\implies$ (1) cont.

$$\prod_{j=1}^{\ell}\left(\frac{c_j}{v_j/\mathbb{1}^T v}\right)^{v_j/\mathbb{1}^T v} = \exp(-D(\frac{v}{\mathbb{1}^T v}, c))$$

$\because$ Definition of relative entropy

$$\geq -\xi\left(D(\frac{v}{\mathbb{1}^T v}, c) + \log(\xi) - 1\right) \;\forall\, \xi > 0$$

$\because$ The convex-conjugate entropy is the negative exponential, i.e., $\exp(-\rho) = \sup_{\xi\in\mathbb{R}_+} -\xi[\rho + \log(\xi) - 1]$. [1]

---

[1] the reasoning behind this uses the Legendre transform but is apparently well know in the field as it is stated without source in the paper.

Proof

(2) $\implies$ (1) cont.

$$-\xi\big(D(\frac{v}{\mathbb{1}^T v}, c) + \log(\xi) - \log(\exp(1))\big)$$

$$= -\xi\Big(\Big(\sum_{j=1}^{n} \frac{v_j}{\mathbb{1}^T v}\log(\frac{v_j}{c_j \mathbb{1}^T v}) + \log(\xi/\exp(1))\Big)$$

$$= -\sum_{j=1}^{n} \frac{\xi v_j}{\mathbb{1}^T v}\log(\frac{\xi v_j}{\mathbb{1}^T v} \cdot \frac{1}{\exp(1) c_j}) = -D(\frac{\xi v}{\mathbb{1}^T v}, \exp(1)c)$$

$\because$ Definition of relative entropy

$$\geq D(v, \exp(1)c) \geq -b$$

$\because$ Choosing $\xi = \mathbb{1}^T v$ and (2.1)

Proof

(1) $\implies$ (2)

$$
\begin{aligned}
& g(x) \geq 0 \\
\iff & g(x)\exp(-\alpha^T x) \geq 0 \\
\iff & \sum_{j=1}^{\ell} c_j \exp((\alpha^{(j)^T} - \alpha^T)x) \geq -b
\end{aligned}
$$

$\because \exp(-\alpha^T x) > 0$ and definition of $g(x)$

Now consider the following lower bound:

$$
\begin{aligned}
& \inf_{x\in\mathbb{R}^n} c^T t \\
& \text{s.t.}\ \exp((\alpha^{(j)^T} - \alpha^T)x) = t_j\ \forall\, j \in [\ell]
\end{aligned}
$$

Proof

(1) $\implies$ (2) cont.

Which is equivelent to the bound of

$$\begin{aligned} p = \inf_{x \in \mathbb{R}^n} & c^T t \\ \text{s.t.} & \exp(({\alpha^{(j)}}^T - \alpha^T)x) \leq t_j \; \forall j \in [\ell] \end{aligned}$$

$\because$ the c_j's are positive.

The Lagrangian dual of this programme (modulo quite a few steps) is

$$\begin{aligned} d = \sup_{v \in \mathbb{R}^\ell_+} & -D(v, \exp(1)c) \\ \text{s.t.} & \sum_{j=1}^{\ell} v_j \alpha^{(j)} = (\mathbb{1}^T v)\alpha \end{aligned}$$

Proof

(1) $\implies$ (2) cont.

- By assumption, $g(x) \geq 0$, we have strict primal feasibility.
- By Slater's condition strong duality holds, i.e., $d = p \geq -b$.
- Hence there exists a $v \in \mathbb{R}_+^\ell$ such that $D(v, \exp(1)c) \leq b$ and $\sum_{j=1}^{\ell} v_j \alpha^{(j)} = (\mathbb{1}^T v)\alpha$.

$\square$

Example

For fixed $\lambda \in [-\frac{1}{2}, \frac{1}{2}]$ consider

$$g_\lambda(x) = \exp(x_1) + \exp(x_2) - \exp(\underbrace{(\frac{1}{2} + \lambda)x_1 + (\frac{1}{2} - \lambda)x_2}_{\alpha^T x})$$

By the lemma we have that $g_\lambda(x) \geq 0$ because
$v := [\frac{1}{2} + \lambda, \frac{1}{2} - \lambda]^T = \alpha \geq 0$ satisfies

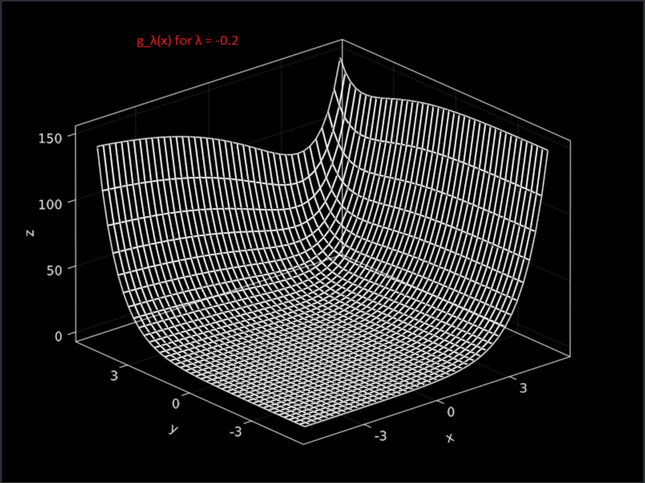
$$D(v, \exp(1)c) - b$$

$$= (\frac{1}{2} + \lambda)\log(\frac{1 + 2\lambda}{2\exp(1)}) + (\frac{1}{2} - \lambda)\log(\frac{1 - 2\lambda}{2\exp(1)}) + 1 \leq 0$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} v = (\mathbb{1}^T v)\alpha = \alpha$$

Example

Visualization of the example



Outline for Section 2

Idea

Sums of AM-GM-Exponentials

- Use AM-GM-Exponentials as the "atoms of positivity".
- Define the cone of Sums of AM-GM-Exponentials (SAGE).
- Recast SAGE into *(second order) Relative Entropy cone*.
- Characterize the *Relative Entropy cone*, $\mathrm{C}_{\mathrm{SAGE}}$.

Definitions

SAGE & $\mathrm{C}_{\mathrm{SAGE}}$

Definition

Given a finite collection of vectors $M \subset \mathbb{R}^n$ the

$$\mathrm{SAGE}(M) := \left\{ f \mid f = \sum_{i=1}^{m} f_i,\ f_i\text{'s are AM/GM-expo.'s in M} \right\}$$

Definition

Suppose $M = \{\alpha^{(j)}\}_{j=1}^{\ell}$ then

$$\mathrm{C}_{\mathrm{SAGE}}(\{\alpha^{(j)}\}_{j=1}^{\ell}) := \left\{ c \in \mathbb{R}^{\ell} \mid \sum_{j=1}^{\ell} c_j \exp(\alpha^{(j)^T} x) \in \mathrm{SAGE}(M) \right\}$$

Counter-Example

A nonnegative non-member of SAGE

$f(x_1, x_2, x_3) = (\exp(x_1) - \exp(x_2) - \exp(x_3))^2$ is non negative but not SAGE.

Why? Start by multiplying out: $f(x_1, x_2, x_3) =$

$$\sum_{i=1}^{3} e^{2e_i^T x} + 2e^{[0,1,1]x} - 2e^{[1,1,0]x} - 2e^{[1,0,1]x}$$

Coefficients: $(1, 1, 1, 2)$ and $-2, -2$

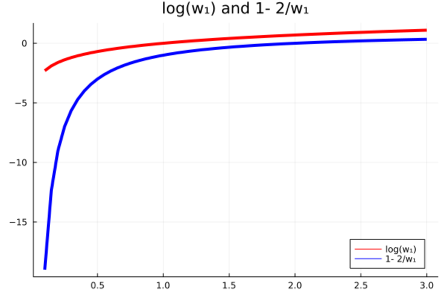
Exponents: $\begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$

Counter-Example cont.

By the Lemma :

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} w = (w_1 + w_2) \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

$$\implies w_1 = 0$$



$$D(w, exp(1)[1,1]^T)$$

$$= w_2 \log(\frac{w_2}{exp(1)}) \leq -2$$

$$\implies \log(w_2) \leq 1 - \frac{2}{w_2}$$

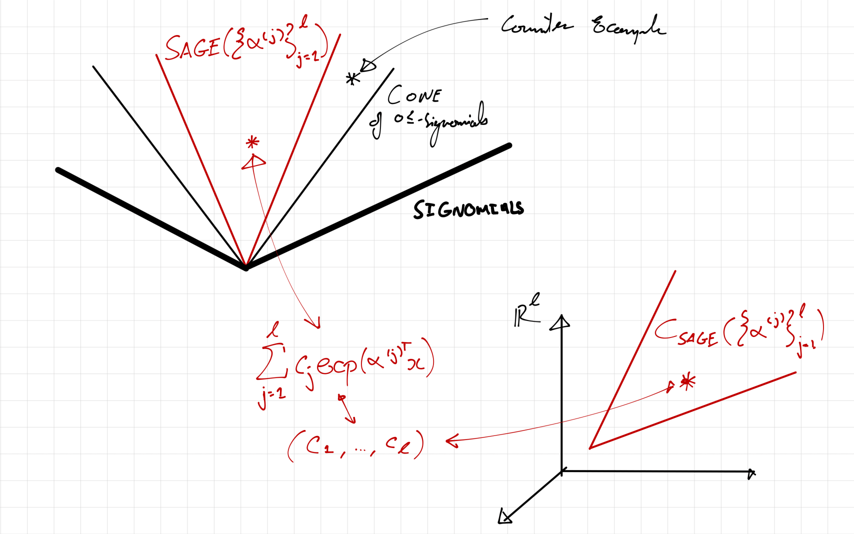
Proposition

$\mathrm{C}_{\mathrm{SAGE}}$ *characterization*

Proposition

$$\begin{aligned}\mathrm{C}_{\mathrm{SAGE}}(\{\alpha^{(j)}\}_{j=1}^{\ell}) = \Big\{ c \in \mathbb{R}^{\ell} \mid \exists c^{(j)} \in \mathbb{R}^{\ell},\ v^{(j)} \in \mathbb{R}^{\ell} \\ \text{s.t. } c = \sum_{j=1}^{\ell} c_j, \\ \sum_{i=1}^{\ell} \alpha^{(i)} v_i^{(j)} = 0,\ v_j^{(j)} = -\mathbb{1}^T v_{\backslash j}^{(j)}, \\ D(v_{\backslash j}^{(j)}, \exp(1) c_{\backslash j}^{(j)}) - c_j^{(j)} \leq 0,\ v_{\backslash j}^{(j)} \in \mathbb{R}_{+}^{\ell-1} \Big\}\end{aligned}$$

Rough schema



Outline for Section 3

1. AM-GM-exponentials and relative entropy
    1.1 Definitions
    1.2 Lemma
    1.3 Example
2. The SAGE Cone
    2.1 Definitions
    2.2 Proposition
3. Signomial Optimization
    3.1 Unconstrained Signomial Optimization
    3.2 Constrained Signomial Optimization
4. Conclusion

Signomial Optimization (Unconstrained)

$$
\begin{aligned}
f_* = \inf_{x \in \mathbb{R}^n} f(x) = f_* = \sup_{\gamma \in \mathbb{R}} &\ \gamma \\
\text{s.t. } &\ f(x) - \gamma \geq 0
\end{aligned}
$$

Assume w.l.o.g. that $f(x) = \sum_{j=1}^{\ell} c_j \exp(\alpha^{(j)^T} x)$ with $\alpha^{(1)} = 0$.
Relax the nonnegativity to SAGE membership

$$
\begin{aligned}
f_{\mathrm{SAGE}} = \sup_{\gamma \in \mathbb{R}} &\ \gamma \\
\text{s.t. } &\ f(x) - \gamma \in \mathrm{SAGE}(\{\alpha^{(j)}\}_{j=1}^{\ell}) \\
&\iff (\because \mathrm{Lemma}) \\
\text{s.t. } &\ (c_1 - \gamma, c_2, \ldots, c_\ell) \in \mathrm{C}_{\mathrm{SAGE}}(\{0, \alpha^{(2)}, \ldots, \alpha^{(\ell)}\})
\end{aligned}
$$

SAGE Optimization (Unconstrained)

Observations

$$
\begin{aligned}
f_{\mathrm{SAGE}} = \sup_{\gamma \in \mathbb{R}} &\ \gamma \\
\text{s.t. } &(c_1 - \gamma, c_2, \ldots, c_\ell) \in \mathrm{C}_{\mathrm{SAGE}}(\{0, \alpha^{(2)}, \ldots, \alpha^{(\ell)}\})
\end{aligned}
$$

- We can solve these problems using second order cone programming (SOCP).
- There is software:
  - CVXPY (modelling in Python)
  - ECOS (solver)

Signomial Optimization (Unconstrained)

A "Pólya type" hierarchy

Definition

$$f^{(p)}_{\mathrm{SAGE}} = \sup_{\gamma \in \mathbb{R}} \gamma$$

$$\text{s.t.} \left(\sum_{j=1}^{\ell} \exp(\alpha^{(j)^T} x) \right)^p \Big(f(x) - \gamma \Big) \in \mathrm{SAGE}\Big(E_{p+1}\big(\{ \alpha^{(j)} \}_{j=1}^{\ell} \big) \Big)$$

where

$$E_{p+1}\big(\{ \alpha^{(j)} \}_{j=1}^{\ell} \big) := \left\{ \sum_{j=1}^{\ell} \lambda_j \alpha^{(j)} \mid \sum_{j=1}^{\ell} \lambda_j \le p, \ \lambda_j \in \mathbb{N} \right\}$$

Signomial Optimization (Unconstrained)

Observations

- $f^{(0)}_{\mathrm{SAGE}} = f_{\mathrm{SAGE}} \because E_1\left(\{\alpha^{(j)}\}_{j=1}^{\ell}\right) = \{\alpha^{(j)}\}_{j=1}^{\ell}$
- $f^{(p)}_{\mathrm{SAGE}} \leq f_* \because$ A feasible solution to $f^{(p)}_{\mathrm{SAGE}}$ would imply $f(x) - \gamma \geq 0$ since the multiplier is positive.
- $f^{(p)}_{\mathrm{SAGE}} \leq f^{(p+1)}_{\mathrm{SAGE}} \because E_{p+1}\left(\{\alpha^{(j)}\}_{j=1}^{\ell}\right) \subset E_{p+2}\left(\{\alpha^{(j)}\}_{j=1}^{\ell}\right)$
- $f^{(p)}_{\mathrm{SAGE}} \to f_*$? as $p \to \infty$ To be proven in the sequel to this talk.

Signomial Optimization (Constrained)

Definitions

Let $f(x), g_1(x), \ldots, g_m(x)$ be signomials in exponents $\{\alpha^{(j)}\}_{j=1}^{\ell}$.
Define

$$C := \{g_j(x) \mid j \in [m]\}$$

and

$$K_C := \{x \in \mathbb{R}^n \mid g(x) \geq 0 \forall g \in C\}$$

$$f_* = \inf_{x \in K} f(x) = f_* = \sup_{\gamma \in \mathbb{R}} \gamma \qquad \text{(C-SP)}$$

$$\text{s.t. } f(x) - \gamma \geq 0 \ \forall \ x \in K_C$$

Signomial Optimization (Constrained)

Attack from weak duality

$$f_{WD} = \sup_{\gamma \in \mathbb{R},\ \mu \in \mathbb{R}^m_+} \gamma$$

$$\text{s.t. } f(x) - \gamma - \sum_{j=1}^{m} \mu_j g_j(x) \geq 0$$

However we are still stuck with two problems:

1. It is still intractable to check ≥ 0 for general signomials.
2. The problem (C-SP) is non-convex in
   - Hence, we do not expect strong duality to hold
     - So, generally $f_{WD} < f_*$

Signomial Optimization (Constrained)

Add "implied" constraints

Definition

$R_q(C) := \left\{ \prod_{k=1}^{q} h_k \mid h_k \in \{1\} \cup C \right\}$

Definition

$$f_{\mathrm{SAGE}}^{(p,q)} = \sup_{\gamma \in \mathbb{R}, s_h \in \mathrm{SAGE}(E_p(\{\alpha^{(j)}\}_{j=1}^{\ell}))} \gamma$$

$$\text{s.t. } f(x) - \gamma - \sum_{h(x) \in R_q(C)} s_h(x) h(x) \in \mathrm{SAGE}(E_{p+q}(\{\alpha^{(j)}\}_{j=1}^{\ell}))$$

Outline for Section 4

Conclusion

- Basically SOS with squares replaced by AM-GM-expo.'s.
- Entropy characterization of AM-GM-expo.'s
- Obtain a double hierarchy because the constraints and objective can be relaxed independently of each other.