Epigraphs of separable spectral functions as exotic cones

Lea Kapelevich, with Chris Coey and Juan Pablo Vielma

June 2023

SIAM OPT

- In previous talks we introduced our generic interior point solver Hypatia
- Hypatia allows users to define their own *exotic* cones
- We confine *exotic* cones to those for which we have analytic oracles
- We call cones that other solvers know how to optimize over the *standard* cones

Background

Standard cones:

1. $\mathbb{R}_+$
2. $\mathbb{S}_+$
3. $\{(u, w) : u \geq \|w\|\}$
4. $\{(u, v, w) : 2uv \geq \|w\|^2\}$
5. $\{(u, v, w) : u \geq v \exp(w/v)\}$
6. $\{(u, v, w) : |w| \leq u^\alpha v^{1-\alpha}\}$

Exotic cones:

1. ℓ_∞ / ℓ_1-norm cones
2. Generalized power cone
3. Sum-of-squares polynomials
4. Sparse PSD matrices
5. LMI cone

...23 in Hypatia

Background

Standard cones:

1. $\mathbb{R}_+$
2. $\mathbb{S}_+$
3. $\{(u, w) : u \geq \|w\|\}$
4. $\{(u, v, w) : 2uv \geq \|w\|^2\}$
5. $\{(u, v, w) : u \geq v \exp(w/v)\}$
6. $\{(u, v, w) : |w| \leq u^\alpha v^{1-\alpha}\}$

Exotic cones:

1. ℓ_∞ / ℓ_1-norm cones
2. Generalized power cone
3. Sum-of-squares polynomials
4. Sparse PSD matrices
5. LMI cone

...23 in Hypatia

- More specialized cones → smaller, *natural* formulations that are faster to solve and simpler to write (see *https://arxiv.org/abs/2005.01136*)

A separable spectral function cone

Cones that look like this:

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{>} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

A separable spectral function cone

Cones that look like this:

$$\mathcal{K} = \mathsf{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{>} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

- $(v, W) \mapsto v\varphi(W/v)$ is the *perspective function* of φ

A separable spectral function cone

Cones that look like this:

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{>} \times \mathbb{S}_{+}^{d} : u \geq v\varphi(W/v)\}$$

- $(v, W) \mapsto v\varphi(W/v)$ is the *perspective function* of φ
- φ is a *separable spectral* function and convex on $\mathbb{S}_{+}$

A separable spectral function cone

Cones that look like this:

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{>} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

- $(v, W) \mapsto v\varphi(W/v)$ is the *perspective function* of φ
- φ is a *separable spectral* function and convex on $\mathbb{S}_+$
- For this talk, think of spectral functions as function of eigenvalues of a matrix

A separable spectral function cone

Cones that look like this:

$$\mathcal{K} = \text{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{>} \times \mathbb{S}_{+}^{d} : u \geq v\varphi(W/v)\}$$

- $(v, W) \mapsto v\varphi(W/v)$ is the *perspective function* of φ
- φ is a *separable spectral* function and convex on $\mathbb{S}_{+}$
- For this talk, think of spectral functions as function of eigenvalues of a matrix
- $\varphi(W) = \sum_i h(\lambda_i(W)) = \text{tr}(h(W))$

A separable spectral function cone

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

A separable spectral function cone

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

$h(w) = -\log(w)$

A separable spectral function cone

$$\mathcal{K} = \text{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

$$h(w) = -\log(w) \qquad\qquad \varphi(W) = -\text{logdet}(W)$$

A separable spectral function cone

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

$h(w) = -\log(w)$ $\qquad$ $\varphi(W) = -\operatorname{logdet}(W)$ $\qquad$ $u \geq -v \operatorname{logdet}(W/v)$

A separable spectral function cone

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

$$h(w) = -\log(w) \qquad \varphi(W) = -\operatorname{logdet}(W) \qquad u \geq -v \operatorname{logdet}(W/v)$$

$$h(w) = w\log(w) \qquad \varphi(W) = \operatorname{tr}(W\log(W)) \qquad u \geq \operatorname{tr}(W\log(W/v))$$

A separable spectral function cone

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

$$\begin{array}{lll} h(w) = -\log(w) & \varphi(W) = -\operatorname{logdet}(W) & u \geq -v \operatorname{logdet}(W/v) \\ h(w) = w\log(w) & \varphi(W) = \operatorname{tr}(W\log(W)) & u \geq \operatorname{tr}(W\log(W/v)) \\ h(w) = w^p & \varphi(W) = \operatorname{tr}(W^p) & u \geq \operatorname{tr}(W^p/v^{p-1}) \end{array}$$

Extended formulation

Suppose $\varphi(W) = f(\lambda(W))$. From Ben-Tal and Nemirovski [2, Proposition 4.2.1.]:

$$
\begin{aligned}
(u, v, W) &\in \mathcal{K} \\
&\Updownarrow \\
\exists x_1, \ldots, x_d, v &: \\
u &\geq v f(x/v) \\
x_1 \geq x_2 \geq \ldots &\geq x_d \\
\textstyle\sum_{i=1}^{j} \lambda_i(W) &\leq \textstyle\sum_{i=1}^{j} x_i, \qquad \forall j \in 1, \ldots, d
\end{aligned}
$$

A barrier function

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

- f is *ν-logarithmically homogeneous* if $f(\theta w) = f(w) - \nu \log(\theta)$ for all $\theta \in \mathbb{R}$

A barrier function

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

- f is *ν-logarithmically homogeneous* if $f(\theta w) = f(w) - \nu \log(\theta)$ for all $\theta \in \mathbb{R}$
- The "obvious" guess for a barrier is:

$$\Gamma(u, v, W) = -\log(v) - \mathrm{logdet}(W) - \log(u - v\varphi(W/v))$$

A barrier function

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

- f is *ν-logarithmically homogeneous* if $f(\theta w) = f(w) - \nu \log(\theta)$ for all $\theta \in \mathbb{R}$
- The "obvious" guess for a barrier is:

$$\Gamma(u, v, W) = -\log(v) - \mathrm{logdet}(W) - \log(u - v\varphi(W/v))$$

- Γ is logarithmically homogeneous with parameter $2 + d$, but we also need self-concordance

A barrier function

$$\mathcal{K} = \mathrm{cl}\{(u, v, W) \in \mathbb{R} \times \mathbb{R}_{++} \times \mathbb{S}^d_+ : u \geq v\varphi(W/v)\}$$

- f is *ν-logarithmically homogeneous* if $f(\theta w) = f(w) - \nu \log(\theta)$ for all $\theta \in \mathbb{R}$
- The "obvious" guess for a barrier is:

$$\Gamma(u, v, W) = -\log(v) - \mathrm{logdet}(W) - \log(u - v\varphi(W/v))$$

- Γ is logarithmically homogeneous with parameter $2 + d$, but we also need self-concordance
- $|\nabla^3\Gamma(x)[h, h, h]| \leq 2\nabla^2\Gamma(x)[h, h]^{3/2}$ for all x, h

When is the barrier SC?

Definition
h is *matrix monotone* if $w_1 \succeq w_2$ implies $h(w_1) \succeq h(w_2)$ for all $w_1, w_2 \in \mathbb{S}^d$ for all integers d.

- E.g. $W \mapsto -W^{-1}$ is matrix-monotone
- E.g. $W \mapsto \exp(W)$ is monotone, but not matrix-monotone

When is the barrier SC?

Definition
h is *matrix monotone* if $w_1 \succeq w_2$ implies $h(w_1) \succeq h(w_2)$ for all $w_1, w_2 \in \mathbb{S}^d$ for all integers d.

- E.g. $W \mapsto -W^{-1}$ is matrix-monotone
- E.g. $W \mapsto \exp(W)$ is monotone, but not matrix-monotone

We will now require that $\varphi(W) = \mathrm{tr}(h(W))$ is such that h' is matrix monotone.

When is the barrier SC?

Let:

$$\mathcal{S} = \{(u, W) : u \geq \varphi(W)\}$$

$\mathcal{K}$ is the *conic hull* of $\mathcal{S}$.

Faybusovich and Tsuchiya [4] showed that under the same conditions, $\mathcal{S}$ admits the $(1+d)$-self-concordant barrier:

$$(u, W) \mapsto -\operatorname{logdet}(W) - \log(u - \varphi(W)).$$

Modified result by Nesterov and Nemirovskii [5, Proposition 5.1.4] gives an LHSCB with parameter $9.48(1+d)$ for $\mathcal{K}$.

When is the barrier SC?

Let:

$$\mathcal{S} = \{(u, W) : u \geq \varphi(W)\}$$

$\mathcal{K}$ is the *conic hull* of $\mathcal{S}$.

Faybusovich and Tsuchiya [4] showed that under the same conditions, $\mathcal{S}$ admits the $(1+d)$-self-concordant barrier:

$$(u, W) \mapsto -\operatorname{logdet}(W) - \log(u - \varphi(W)).$$

Modified result by Nesterov and Nemirovskii [5, Proposition 5.1.4] gives an LHSCB with parameter $9.48(1+d)$ for $\mathcal{K}$.

Theorem
Γ *in is a* $(2+d)$*-LHSCB for* $\mathcal{K}$.

Proof.
See *https://pubsonline.informs.org/doi/abs/10.1287/moor.2022.1324*. □

NB: a newer result

Newer result by Fawzi and Saunderson [3] implies

- Γ is self-concordant for matrix-convex h (rather than matrix-monotone h')
- $2 + d$ is the *optimal* barrier parameter

The dual cones

- Adding a cone in Hypatia makes both primal and dual available
- Let:

$$\varphi^*(R) = \sup_{W \succeq 0}\{-\langle W, R \rangle - \varphi(W)\}$$

- Then:

$$\mathcal{K}^* := \text{cl}\{(u, v, W) : u > 0, v \geq u\varphi^*(W/u)\}$$

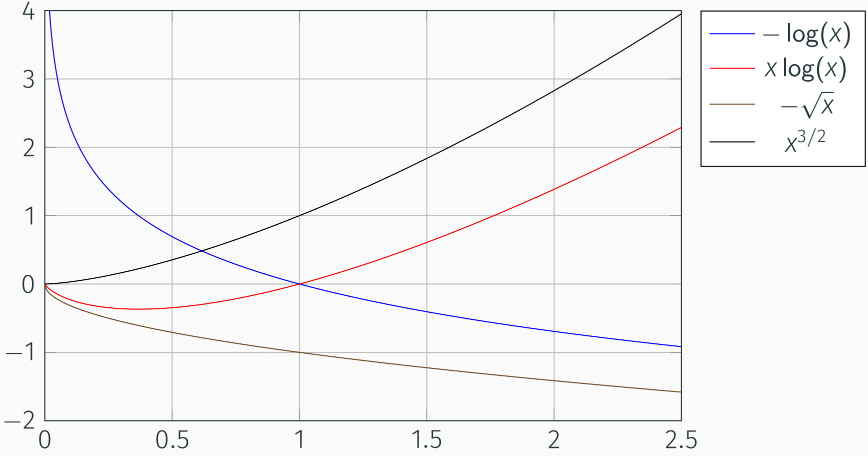
- This is a consequence of Rockafellar [6, Theorem 14.4] and Zhang [7, Theorem 3.2]
- If $\varphi = \text{tr}(h(W))$ then $\varphi^* = \text{tr}(h^*(W))$ due to [1, Lemma 29 and Theorem 30]

Examples

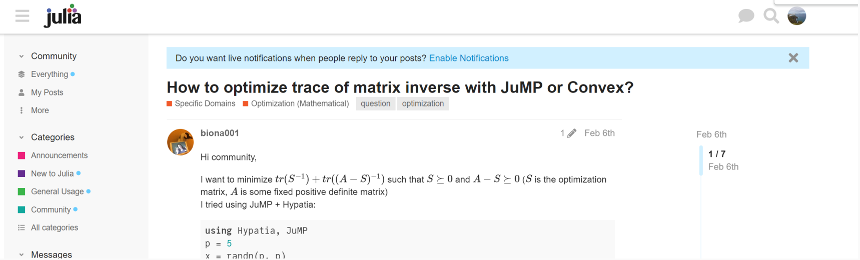
function	h	h'	$dom(h^*)$	h^*
NegLog	$-\log(x)$	$-x^{-1}$	$\mathbb{R}_{++}$	$-1-\log(x)$
NegEntropy	$x\log(x)$	$1+\log(x)$	$\mathbb{R}$	$\exp(-1-x)$
NegSqrt	$-\sqrt{x}$	$-\frac{1}{2}x^{-1/2}$	$\mathbb{R}_{++}$	$\frac{1}{4}x^{-1}$
NegPower, $p \in (0,1)$	$-x^p$	$-px^{p-1}$	$\mathbb{R}_+$	$-(p-1)(x/p)^q$
Power, $p \in (1,2]$	x^p	px^{p-1}	$\mathbb{R}$	$(p-1)(x_-/p)^q$

$q := p/(p-1)$ and $x_- := \max(-x, 0)$

Examples



Hypatia’s generic cone



Hypatia’s generic cone

```julia
import Hypatia
import Hypatia.Cones.EpiPerSepSpectralCone
import Hypatia.Cones.NegSqrtSSF
import Hypatia.Cones.MatrixCSqr
using JuMP

model = Model(Hypatia.Optimizer)
@variable(model, x[1:3])
cone = EpiPerSepSpectralCone{Float64}(NegSqrtSSF(),
    MatrixCSqr{Float64, Float64}, 1, true)
@constraint(model, x in cone)
```

The Hessian

For

$$\Gamma(u, v, W) := -\log(u - v\varphi(W/v)) - \log(v) - \operatorname{logdet}(W),$$

The Hessian

For

$$\Gamma(u, v, W) := -\log(u - v\varphi(W/v)) - \log(v) - \operatorname{logdet}(W),$$

$$\nabla^2\Gamma = \begin{bmatrix} \nabla^2\Gamma_{u,u} & \nabla^2\Gamma_{u,v} & \dots & \nabla^2\Gamma_{u,w} & \dots \\ \nabla^2\Gamma_{v,u} & \nabla^2\Gamma_{v,v} & \dots & \nabla^2\Gamma_{v,w} & \dots \\ \vdots & \vdots & \ddots & & \\ \nabla^2\Gamma_{w,u} & \nabla^2\Gamma_{w,v} & & \nabla^2\Gamma_{w,w} & \\ \vdots & \vdots & & & \ddots \end{bmatrix}$$

The Hessian

$$
\begin{aligned}
\nabla^2 \Gamma_{u,u} &= \zeta^{-2}, \\
\nabla^2 \Gamma_{v,u} &= -\zeta^{-2}\sigma, \\
\nabla^2 \Gamma_{w,u} &= -\zeta^{-2}\nabla\varphi, \\
\nabla^2 \Gamma_{v,v} &= v^{-2} + \zeta^{-2}\sigma^2 + v^{-1}\zeta^{-1}\nabla^2\varphi[w/v, w/v], \\
\nabla^2 \Gamma_{w,v} &= \zeta^{-2}\sigma\nabla\varphi - v^{-1}\zeta^{-1}\nabla^2\varphi[w/v], \\
\nabla^2 \Gamma_{w,w} &= \zeta^{-2}(\nabla\varphi)(\nabla\varphi)' + v^{-1}\zeta^{-1}\nabla^2\varphi + W^{-1} \otimes W^{-1} \\
&= \zeta^{-2}(\nabla\varphi)(\nabla\varphi)' + (G \otimes G)\,\mathrm{Diag}(\mathrm{vec}(M))(G \otimes G)'.
\end{aligned}
$$

where:

- $\zeta = u - v\varphi(W/v)$
- $\sigma = \varphi - \nabla\varphi[w/v]$
- $W = G\,\mathrm{Diag}(\lambda)G'$
- $M_{i,j} = v^{-1}\zeta^{-1}\frac{h'(\lambda_i)-h'(\lambda_j)}{\lambda_i-\lambda_j} + \lambda_i^{-1}\lambda_j^{-1}$

The Hessian

$$
\begin{aligned}
\nabla^2\Gamma_{u,u} &= \zeta^{-2}, \\
\nabla^2\Gamma_{v,u} &= -\zeta^{-2}\sigma, \\
\nabla^2\Gamma_{w,u} &= -\zeta^{-2}\nabla\varphi, \\
\nabla^2\Gamma_{v,v} &= v^{-2} + \zeta^{-2}\sigma^2 + v^{-1}\zeta^{-1}\nabla^2\varphi[w/v, w/v], \\
\nabla^2\Gamma_{w,v} &= \zeta^{-2}\sigma\nabla\varphi - v^{-1}\zeta^{-1}\nabla^2\varphi[w/v], \\
\nabla^2\Gamma_{w,w} &= \zeta^{-2}(\nabla\varphi)(\nabla\varphi)' + v^{-1}\zeta^{-1}\nabla^2\varphi + W^{-1}\otimes W^{-1} \\
&= \zeta^{-2}(\nabla\varphi)(\nabla\varphi)' + (G\otimes G)\operatorname{Diag}(\operatorname{vec}(M))(G\otimes G)'.
\end{aligned}
$$

Note that:

$$
(G\otimes G)\operatorname{vec}(R) = \operatorname{vec}(G'RG)
$$

$$
\left((G\otimes G)\operatorname{Diag}(\operatorname{vec}(M))(G\otimes G)'\right)^{-1} = (G\otimes G)\operatorname{Diag}(\operatorname{vec}(M))^{-1}(G\otimes G)'
$$

The inverse Hessian

$$
\begin{aligned}
(\nabla^2\Gamma)^{-1}_{u,u} &= k_3 k_4(\zeta^2 + \beta),\\
(\nabla^2\Gamma)^{-1}_{v,u} &= k_1 k_4,\\
(\nabla^2\Gamma)^{-1}_{w,u} &= \alpha + k_1 k_4 \gamma,\\
(\nabla^2\Gamma)^{-1}_{v,v} &= k_4,\\
(\nabla^2\Gamma)^{-1}_{w,v} &= k_4 \gamma,\\
(\nabla^2\Gamma)^{-1}_{w,w} &= k_4 \gamma\gamma' + (G \otimes G)\,\mathrm{Diag}(\mathrm{vec}(M))^{-1}(G \otimes G)'.
\end{aligned}
$$

where:

$$
\begin{aligned}
\alpha &:= (G \otimes G)\,\mathrm{Diag}(\mathrm{vec}(M)^{-1} \cdot \nabla\varphi)(G \otimes G)'\\
\beta &:= \langle \alpha, \nabla\varphi \rangle\\
\gamma &:= (G \otimes G)\,\mathrm{Diag}(\mathrm{vec}(M)^{-1} \cdot \nabla^2\varphi[w])(G \otimes G)'
\end{aligned}
$$

TOO

For $\mathrm{T} := -\frac{1}{2}\nabla^3\Gamma[\delta,\delta]$ with $\delta = (p, q, r)$:

$$\mathrm{T}_u = \zeta^{-3}\chi^2 + \tfrac{1}{2}v\zeta^{-2}\nabla^2\varphi[\xi,\xi], \tag{7a}$$

$$\mathrm{T}_v = -\mathrm{T}_u\sigma - \mathrm{tr}(\tau \circ W/v) + \tfrac{1}{2}\zeta^{-1}\nabla^2\varphi[\xi,\xi] + q^2v^{-3}, \tag{7b}$$

$$\mathrm{T}_W = -\mathrm{T}_u\nabla\varphi + \tau + W^{-1}RW^{-1}RW^{-1}. \tag{7c}$$

Where:

$$\xi = v^{-1}(r - qv^{-1}W)$$

$$\tau = \zeta^{-2}(\chi + qv^{-1})\nabla^2\varphi[\xi] - \frac{1}{2}\zeta^{-1}\nabla^3\varphi[\xi,\xi]$$

$$\chi = p - q\sigma - \nabla\varphi[r]$$

Most expensive part is done in $O(d^3)$.

Conclusions

- We have a general class of cones for which the “obvious” barrier is self-concordant
- The “separable spectral function” cone has highly structured oracles
    - Main block in the Hessian is a Kronecker plus a rank one term (or a diagonal plus rank one for vector analogs)
    - Inverse Hessian has the same structure as the Hessian (good news for the dual cones)

Appendices

Constants for the inverse Hessian

$$
\begin{aligned}
k_1 &:= \sigma + (\nabla\varphi)[\gamma] \\
k_2 &:= \frac{k_1}{\zeta^2 + \nabla\varphi[\alpha]} \\
k_3 &:= v^{-2} + \sigma k_1 \sum_i (v^{-1}\lambda_i + k_1\alpha_i - \gamma_i)(\nabla^2 h)_i \lambda_i \\
k_4 &:= (k_3 - k_1 k_2)^{-1}
\end{aligned}
$$

References

[1] Michel Baes. Convexity and differentiability properties of spectral functions and spectral mappings on euclidean jordan algebras. *Linear algebra and its applications*, 422(2-3):664–700, 2007.

[2] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*, volume 2. Siam, 2001.

[3] Hamza Fawzi and James Saunderson. Optimal self-concordant barriers for quantum relative entropies. *arXiv preprint arXiv:2205.04581*, 2022.

[4] Leonid Faybusovich and Takashi Tsuchiya. Matrix monotonicity and self-concordance: how to handle quantum entropy in optimization problems. *Optimization Letters*, 11(8):1513–1526, 2017.

[5] Yurii Nesterov and Arkadii Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13. Siam, 1994.

[6] Ralph Tyrell Rockafellar. *Convex analysis*. Princeton university press, 2015.

[7] Shuzhong Zhang. A new self-dual embedding method for convex programming. *Journal of Global Optimization*, 29(4):479–496, 2004.