Spectral functions and an apology of Jordan Algebras

Bruno F. Lourenço
Institute of Statistical Mathematics, Japan
joint work with Akiko Takeda (U. Tokyo)

December 1st, 2020
WoMBat 2020

Prologue

Figure: Jordan Algebras: the vegemite of our community

How/why did (some) optimizers get interested in Jordan Algebras?

- It was during the frenetic days of the interior-point method (IPM) era.

$$\begin{aligned} \inf_{x} \quad & \langle c, x \rangle \\ \text{subject to} \quad & \mathcal{A}x = b \\ & x \in \mathcal{K} \end{aligned}$$

- Nesterov and Todd showed that "good" **primal-dual** IPMs exists when $\mathcal{K}$ is a "**self-scaled cone**".
- Güler pointed out that **self-scaled cones** coincide with **symmetric cones**.
- Faybusovich developed primal-dual IPMs using **Jordan Algebras**.

What is a Jordan Algebra?

A **Euclidean Jordan Algebra** is a real vector space $\mathcal{E}$ equipped with a bilinear product $\circ : \mathcal{E} \times \mathcal{E} \to \mathcal{E}$ and an inner product $\langle \cdot, \cdot \rangle$, satisfying

1. $y \circ z = z \circ y$,
2. $\langle y \circ z, w \rangle = \langle y, z \circ w \rangle$,
3. $y \circ (y^2 \circ z) = y^2 \circ (y \circ z)$,

There is an identity element e:

$$e \circ x = x \qquad \forall x.$$

The cone

$$\mathcal{K} = \{x \circ x \mid x \in \mathcal{E}\}$$

is **symmetric**, i.e., **homogeneous and self-dual**.

Spectral Decomposition

Let $(\mathcal{E}, \circ)$ be an Euclidean Jordan Algebra.
Let $x \in \mathcal{E}$, then x can be written as

$$x = a_1 c_1 + \cdots + a_r c_r, \qquad \text{(Spectral Decomposition)}$$

where $c_i \circ c_i = c_i$ and $c_i \circ c_j = 0$, for $i \neq j$. $c_1 + \cdots + c_r = e$.

- The a_i are the eigenvalues of x.
- r is the rank of $\mathcal{E}$.
- $[c_1, \ldots, c_r]$ is called a **Jordan frame**.

Examples

$\mathcal{E} = \mathcal{S}^n$ (symmetric matrices). We have

- $X \circ Y = (XY + YX)/2$
- $e = I_n$
- $\mathcal{K} = \mathcal{S}^n_+$ (positive semidefinite matrices).

$\mathcal{E} = \mathbb{R}^n = \mathbb{R} \times \mathbb{R}^{n-1}$. We have

- $x \circ y = (x_1 y_1 + \overline{x}^\top \overline{y}, x_1 \overline{y} + y_1 \overline{x})$
- $e = (1, 0, \dots, 0)$
- $\mathcal{K} = \mathcal{L}^n_2 = \{(x_1, \overline{x}) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid x_1 \geq \|\overline{x}\|_2\}$ (second order cone).

$\mathcal{E} = \mathbb{R}^n$. We have

- $x \circ y = (x_1 y_1, \cdots, x_n y_n)$
- $e = (1, \dots, 1)$
- $\mathcal{K} = \mathbb{R}^n_+$ (nonnegative orthant).

The classification of symmetric cones

$\mathcal{K}$ is **simple** $\iff$ $\mathcal{K}$ cannot be written as $\mathcal{K}^1 \times \mathcal{K}^2$ in a non-trivial way.

1. There are only **5** types of simple symmetric cones (**up to isomorphism**).
   1. Second order cone for $n \geq 3$
   2. $n \times n$ PSD matrices over the reals, complex numbers, quaternions (!).
   3. 3×3 PSD matrices over the octonions (!!!).
2. Every symmetric cone is a direct product of simple symmetric cones.

Criticism #1 - "If there are only 5 of them, what is the point of the Jordan Algebra framework?"

Even if you care only about (real) SDPs and SOCPs, I'd argue that there is still value in being able to prove results about them in an unified manner. But...

The usual arguments:

- It is already the default framework for second order cone programming. See Fukushima, Luo and Tseng ('01); Alizadeh and Goldfarb ('03).
- Provides insight and extra tools even for the already well-understood case of real symmetric cases.
- Who knows? Maybe someone will find a use for quaternion matrices[1] or the 3×3 octonions.

The not so usual argument:

Jordan Algebras give you tools to deal with **block structure** in a natural way.

[1] see *A Low Rank Quaternion Decomposition Algorithm and Its Application in Color Image Inpainting* by Chen, Qi, Zhang and Xu

For solving large scale matrix problems...

$\mathcal{S}^n$: $n \times n$ symmetric matrices with n **big**

$$\begin{aligned} \inf_x \quad & f(x) \\ \text{subject to} \quad & \operatorname{rank}(x) \leq k \\ & x \in \mathcal{S}^n \end{aligned}$$

very little hope of solving if the problem is dense or does not have structure.
Much better to solve this:

$$\begin{aligned} \inf_x \quad & f(x) \\ \text{subject to} \quad & \operatorname{rank}(x) \leq k \\ & x \in \mathcal{S}^{\mathsf{n}_1} \times \cdots \times \mathcal{S}^{\mathsf{n}_\ell}. \end{aligned}$$

with $n_1 + \cdots + n_\ell = n$.
Optimizers prefer to fight 100 duck-sized horses than 1 horse-sized duck.

Some spectral analysis

λ:eigenvalue function, ordered from largest to smallest.
λ_i: i-th largest eigenvalue function

$$\mathrm{rank}\,(x) \leq k \Leftrightarrow \lambda_{n-k+1}(x) = \cdots = \lambda_n(x) = 0 \Leftrightarrow \|\lambda(x)\|_1 - \|\lambda(x)\|_{n-k} = 0$$

σ_i: maps u to its i-th largest element.
Let $x \in \mathcal{S}^n$.

$$\partial\lambda_i(x) = \{Q\mathrm{diag}\,(\mu)Q^T \mid Q : \text{orthogonal}, QxQ^T = \lambda(x), \mu \in \partial\sigma_i(\lambda(x))\}$$

A. S. Lewis.
Nonsmooth analysis of eigenvalues.
Mathematical Programming, 84(1):1–24, Jan 1999.

Some spectral analysis (continued)

Let $x \in \mathcal{S}^n$.

$$\partial\lambda_i(x) = \{Q\mathrm{diag}\,(\mu)Q^T \mid Q \in O(n), QxQ^T = \lambda(x), \mu \in \partial\sigma_i(\lambda(x))\}$$

How about if $x \in \mathcal{S}^1 \times \mathcal{S}^2$?

$$\partial\lambda_i(x) \neq \{Q\mathrm{diag}\,(\mu)Q^T \mid Q \in O(1)\times O(2), QxQ^T = \lambda(x), \mu \in \partial\sigma_i(\lambda(x))\}$$

$QxQ^T = \lambda(x)$ **might not exist**. E.g.,

For $x = (1) \times \begin{pmatrix} 2 & \\ & 3 \end{pmatrix}$, $\nexists\ Q \in O(1) \times O(2)$, such that

$QxQ^T = (3) \times \begin{pmatrix} 2 & \\ & 1 \end{pmatrix}$

Jordan Algebras provides tools to treat $\mathcal{S}^{n_1} \times \mathcal{S}^{n_2}$ with the respect it deserves.

Spectral functions

Let $(\mathcal{E}, \circ)$ be a Jordan Algebra of rank r.
$f : \mathbb{R}^r \to \mathbb{R}$ is permutation invariant iff

$$f(u) = f(Pu),$$

for every permutation matrix P.

Definition (Spectral functions)

$F : \mathcal{E} \to \mathbb{R}$ is spectral if there is a permutation invariant function f such that

$$F(x) = f(\lambda(x)), \quad \forall x \in \mathcal{E}$$

1. $F_1(x) := \|x\|_p$
2. $F_2(x) := -\log\det x$
3. $F_3(x) := \lambda_k(x)$
4. $F_4(x) := \mathrm{rank}\,(x)$

1. $f_1(u) := \|u\|_p$
2. $f_2(u) := -\sum_i \log u_i$
3. $f_3(u) := \max_k(u)$
4. $f_4(u) := |u|_0$

What is known about spectral functions?

$$F(x) := f(\lambda(x))$$

If f is

1. differentiable or twice differentiable, or
2. continuous or Lipschitz continuous, or
3. convex, or
4. quasi-convex, or
5. semismooth or strongly semismooth

then the same is true of F.

M. Baes.
Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras.
Linear Algebra and its Applications, 422(2):664 – 700, 2007.

D. Sun and J. Sun.
Löwner's operator and spectral functions in Euclidean Jordan algebras.
Mathematics of Operations Research, 33(2):421–445, 2008.

Our goal

$$F(x) := f(\lambda(x))$$

Goal

Relate the generalized subdifferentials of f and F.

Regular, approximate and horizon subdifferential

d is a *regular subgradient* of f at u if

$$\liminf_{\substack{v\to 0\\ v\neq 0}} \frac{f(u+v)-f(u)-\langle d,v\rangle}{\|v\|} \geq 0.$$

$\hat{\partial} f(u)$: *regular subdifferential* of f at u.

d is an *approximate subgradient* of f at u if there are $\{u^k\}$, $\{d^k\}$, with $d^k \in \hat{\partial} f(u^k)$ and

$$u^k \to u, \qquad f(u^k) \to f(u), \qquad d^k \to d.$$

$\partial f(u)$: *approximate subdifferential* of f at u.

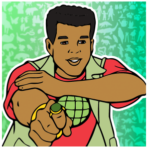
d is an *horizon subgradient* of f at u if there are $\{u^k\}$, $\{d^k\}$, $\{t^k\}$, with $d^k \in \partial^\infty f(u^k)$ and

$$u^k \to u, \qquad f(u^k) \to f(u), \qquad t^k d^k \to d, \qquad t^k \downarrow 0.$$

$\partial^\infty f(u)$: *horizon subdifferential* of f at u.

Some tools...

Tools

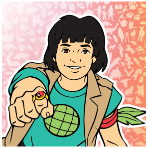


Earth

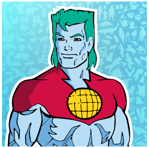
Fire

Wind

Water

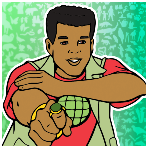


Heart



Captain Planet

Tools (a.k.a. shoulder of giants and all that stuff)

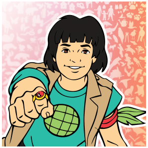


Commutation

Baes' formula

Majorization

Lewis' argument



Jordan Algebra

Theorem (L. and Takeda)

Let $F = f \circ \lambda$.

$$\diamond F(x) = \{s \in \mathcal{E} \mid \exists \mathcal{J} \in \mathcal{J}(x,s) \text{ with } \operatorname{diag}(s,\mathcal{J}) \in \diamond f(\lambda(x))\},$$

where $\diamond$ *is* $\hat{\partial}$, ∂, ∂^∞.

Operator commutation

$x, y \in \mathcal{E}$ **operator commute** in $\mathcal{E}$ if there is a Jordan frame $\mathcal{J} = [c_1, \ldots, c_r]$ such that

$$x = \sum_{i=1}^{r} a_i c_i, \qquad y = \sum_{i=1}^{r} b_i c_i.$$

In other words, x, y can be **simultaneously diagonalized**.
Note: We can always reorder the $[c_1, \ldots, c_r]$ in a way that

$$x = \sum_{i=1}^{r} \lambda_i(x) \hat{c}_i, \qquad y = \sum_{i=1}^{r} \hat{b}_i \hat{c}_i,$$

The set of all such frames will be denote by $\mathcal{J}(x, y)$.

Commutation Principles

Theorem (Ramírez, Seeger and Sossa)

1. $F : \mathcal{E} \to \mathbb{R}$: *spectral function*
2. Θ: *Fréchet differentiable*

If a is a local minimizer/maximizer of the map

$$x \mapsto \Theta(x) + F(x)$$

then a and $\Theta'(a)$ operator commute *in* $\mathcal{E}$.

H. Ramírez, A. Seeger, and D. Sossa.
Commutation Principle for Variational Problems on Euclidean Jordan algebras.
SIAM Journal on Optimization, 23(2):687–694, 2013.

M. Gowda and J. Jeong.
Commutation principles in Euclidean Jordan algebras and normal decomposition systems.
SIAM Journal on Optimization, 27(3):1390–1402, 2017.

Commutation principles and generalized subdifferentials

Proposition (L. and Takeda)

Let $F : \mathcal{E} \to \mathbb{R}$ *be a spectral function. If*

$$s \in \diamond F(x),$$

where $\diamond$ *is* $\hat{\partial}$, ∂ *or* ∂^∞, *then* x *and* s operator commute *in* $\mathcal{E}$.

The diagonal map

Suppose that $\mathcal{J} = [c_1, \dots, c_r]$ is a Jordan frame and

$$x = a_1 c_1 + \cdots a_r c_r.$$

For every i, we have

$$\langle c_i, x \rangle = a_i.$$

We will use the map $\operatorname{diag}(\cdot, \mathcal{J}) : \mathcal{E} \to \mathbb{R}^r$ such that

$$\operatorname{diag}(z, \mathcal{J}) = (\langle c_1, z \rangle, \dots, \langle c_r, z \rangle).$$

Main result

$\mathcal{J} \in \mathcal{J}(x, s)$ is a frame that simultaneously diagonalize x and s and correctly orders the eigenvalues of x.

Theorem (L. and Takeda)

Let $F : \mathcal{E} \to \mathbb{R}$ be a spectral function $F = f \circ \lambda$. For $x \in \mathcal{E}$, we have

$$\diamond F(x) = \{s \in \mathcal{E} \mid \exists \mathcal{J} \in \mathcal{J}(x, s) \text{ with } \operatorname{diag}(s, \mathcal{J}) \in \diamond f(\lambda(x))\},$$

where $\diamond$ is $\hat{\partial}$, ∂ or ∂^{∞}.

If f or F are locally lower semicontinuous, then the result holds for the Clarke subdifferential too.

Generalized subdifferentials of $\lambda_k(\cdot)$

$\mathcal{I}$: set of primitive idempotents of the algebra $\mathcal{E}$.

Theorem (Generalized subdifferentials of $\lambda_k(\cdot)$, L. and Takeda)

$\lambda_k(x)$: k-th largest eigenvalue of x.

$$\partial_C \lambda_k(x) = \operatorname{conv}\{c \in \mathcal{I} \mid x \circ c = \lambda_k(x)c\},$$

$$\hat{\partial}\lambda_k(x) = \begin{cases} \partial_C \lambda_k(x) & \text{if } k = 1 \text{ or } \lambda_{k-1}(x) > \lambda_k(x) \\ \emptyset, & \text{otherwise} \end{cases}$$

$$\partial^\infty \lambda_k(x) = \{0\},$$

$$\partial \lambda_k(x) = \{s \in \partial_C \lambda_k(x) \mid \operatorname{rank} x \leq \alpha\},$$

where $\alpha = 1 - k + |\{i \mid \lambda_i(x) \geq \lambda_k(x)\}|$.

Conclusion

- Other results: F and f share the same Kurdyka–Łojasiewicz exponent, some linear algebraic results for EJAs and etc.
- Next goal: Use the results for the analysis of algorithms!

$$\min_{x \in \mathcal{E}} \quad \Phi(x) = \psi(x) + F(x). \tag{OPT}$$

If x^* is local optimal, then

$$-\nabla\psi(x^*) \in \partial F(x^*).$$

L. and Akiko Takeda
Generalized subdifferentials of spectral functions over Euclidean Jordan algebras
(https://arxiv.org/abs/1902.05270). To appear in SIOPT.

Where to learn EJAs?

M. Baes.
Spectral functions and smoothing techniques on Jordan algebras.
PhD thesis, Université Catholique de Louvain, 2006.

L. Faybusovich.
Several Jordan-algebraic aspects of optimization.
Optimization, 57(3):379–393, 2008.

M. S. Gowda.
Positive and doubly stochastic maps, and majorization in Euclidean Jordan algebras.
Linear Algebra and its Applications, 528:40 – 61, 2017.

Jos F. Sturm.
Similarity and other spectral relations for symmetric cones,
Linear Algebra and its Applications, Volume 312, Issues 1–3, 2000.

Symmetric cones can be hidden in plain sight

Let Q be a matrix with exactly one negative eigenvalue with unit eigenvector q

$$\mathcal{K} = \{x \mid x^T Qx \leq 0, q^T x \geq 0\}$$

$\mathcal{K}$ is a symmetric cone isomorphic to the second order cone.

Symmetric cones can be hidden in plain sight

- $\mathcal{K}$ is homogeneous $\iff$ the automorphism group of $\mathcal{K}$ acts transitively on $\operatorname{int}\mathcal{K}$.
- $\mathcal{K}$ is self-dual $\iff$ $\exists\langle\cdot,\cdot\rangle$ such that $\mathcal{K} = \mathcal{K}^*$, where $\mathcal{K}^* = \{y \mid \langle x, y\rangle \geq 0, \forall x \in \mathcal{K}\}$.
- Could $\mathcal{K}_p = \{(t, x) \mid t \geq \|x\|_p\}$ be homogeneous? Or self-dual? Or symmetric?

M. Ito and L.,
The p-cones in dimension $n \geq 3$ are not homogeneous when $p \neq 2$.
Linear Algebra and its Applications, 2017

M. Ito and L.
The automorphism group and the non-self-duality of p-cones.
Journal of Mathematical Analysis and Applications, 2019.

M. Orlitzky
On the symmetry of induced norm cones.
Optimization, 2020.

Criticism #2 - "Why don't you prove results over larger frameworks?"

Answer: Goldilocks' rule. EJAs are "just right".

- Homogeneous cones: T-algebras
- Hyperbolicity cones: hyperbolic polynomials.

Slightly more honest answer: I would, if I could.

Results for EJAs already used almost 20 years of research developments. The technology is simply not there for hyperbolic polynomials.

H. H. Bauschke, O. Güler, A. S. Lewis, and H. S. Sendov.
Hyperbolic polynomials and convex analysis.
Canadian Journal of Mathematics, 2001.

Project-and-rescale methods and Chubanov's algorithm

- They have an auxiliary method that performs well if some conditioning measure is "good enough".
- If the auxiliary method fails, the problem is rescaled so that the auxiliary method can have another try.

J. Peña and N. Soheili.
Solving Conic Systems via Projection and Rescaling
Mathematical Programming, 2017.

L., T. Kitahara, M. Muramatsu, T. Tsuchiya
An extension of Chubanov's algorithm to symmetric cones.
Mathematical Programming, 2019.

The Jordan product and PSD matrices

Lemma (L.,Fukuda,Fukushima,2016)

Let $\lambda \in \mathcal{S}^m$. *The following are equivalent:*

1. $\lambda \succeq 0$,
2. $\exists y \in \mathcal{S}^m$ *such that* $y \circ \lambda = 0$ *and*

$$\langle w \circ w, \lambda \rangle > 0, \tag{1}$$

for all nonzero $w \in \mathcal{S}^m$ *with* $y \circ w = 0$.

Any y *satisfying* (1) *also satisfies*

$$\operatorname{rank} \lambda = m - \operatorname{rank} y.$$

Moreover, if σ *and* σ' *are nonzero eigenvalues of* y, *then* $\sigma + \sigma' \neq 0$.

L., E. H. Fukuda and M. Fukushima.
Optimality conditions for nonlinear semidefinite programming via squared slack variables
Mathematical Programming, 2018

Extension to symmetric cones

Lemma (L.,Fukuda,Fukushima, 2017)

Let $\lambda \in \mathcal{E}$. The following are equivalent:

1. $\lambda \in \mathcal{K}$,
2. $\exists y \in \mathcal{E}$ *such that*
$$y \circ \lambda = 0 \text{ and } \langle w \circ w, \lambda \rangle > 0, \quad (2)$$
for all $w \in \mathcal{E}$ *with* $y \circ w = 0$.

Any y *satisfying* (2) *satisfies*

$$\operatorname{rank} \lambda = m - \operatorname{rank} y.$$

If σ *and* σ' *are nonzero eigenvalues of* the same block *of* y, *then* $\sigma + \sigma' \neq 0$.

L., E. H. Fukuda and M. Fukushima.
Optimality conditions for problems over symmetric cones and a simple augmented Lagrangian method
Mathematics of Operations Research, 2018

Majorization

Let $u, v \in \mathbb{R}^r$.

$$u \prec v \Leftrightarrow \quad u_1 + \cdots + u_r = v_1 + \cdots + v_r \text{ and } \sum_{i=1}^{k} u_i^{\downarrow} \leq \sum_{i=1}^{k} v_i^{\downarrow}, \ \forall k, 1 \leq k \leq r-1$$

Proposition (Gowda'17)

Let $\mathcal{J}$ be a Jordan frame and let $x \in \mathcal{E}$. Then,

$$\operatorname{diag}(x, \mathcal{J}) \prec \lambda(x).$$

Directional derivatives

Theorem (L. and Takeda)

Let $x, z \in \mathcal{E}$ *and let* $\mathcal{J} \in \mathcal{J}(x)$*. Then*

$$\operatorname{diag}(z, \mathcal{J}) \in \operatorname{conv}\{P\lambda'(x; z) \mid P \in \mathcal{P}^r(\lambda(x))\}$$

$\mathcal{P}^r(\lambda(x))$: permutations that fix $\lambda(x)$.

Majorization principles

Let $u \in \mathbb{R}^r$, we denote by $u^\downarrow$ the reordering of the coordinates of u in such a way that

$$u_1^\downarrow \geq \cdots \geq u_r^\downarrow.$$

Let $v \in \mathbb{R}^r$. We say that u is *majorized* by v ($u \prec v$) if

$$\sum_{i=1}^{k} u_i^\downarrow \leq \sum_{i=1}^{k} v_i^\downarrow$$

holds for every $k \in \{1, \ldots, r-1\}$ and

$$\sum_{i=1}^{r} u_i^\downarrow = \sum_{i=1}^{r} v_i^\downarrow.$$

If $x, y \in \mathcal{E}$ we say that x *is majorized by* y ($x \prec y$) if $\lambda(x) \prec \lambda(y)$.

Majorization principles

Proposition (Classical linear algebra)

Let,$u, v \in \mathbb{R}^r$. Then, $u \prec v$ if and only if there exists a doubly stochastic matrix A such that $u = Av$. From Birkhof's theorem, $u \prec v$ if and only if

$$u \in \operatorname{conv}\{Pv \mid P \in \mathcal{P}^r\}.$$

Proposition (Gowda)

Let $\mathcal{J}$ be a Jordan frame and let $x \in \mathcal{E}$. Then, $\operatorname{diag}(x, \mathcal{J})$ is majorized by $\lambda(x)$. In particular,

$$\operatorname{diag}(x, \mathcal{J}) \in \operatorname{conv}\{P\lambda(x) \mid P \in \mathcal{P}^r\}.$$

$\mathcal{P}^r$: $r \times r$ permutation matrices.

M. S. Gowda.
Positive and doubly stochastic maps, and majorization in Euclidean Jordan algebras.
Linear Algebra and its Applications, 528:40 – 61, 2017.