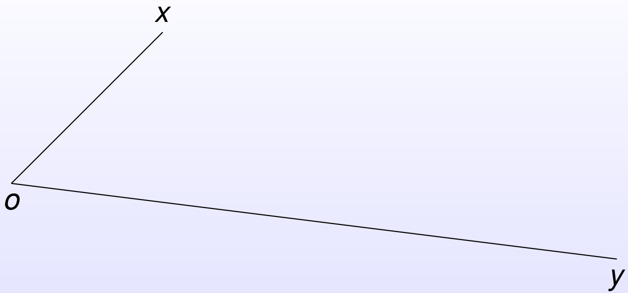
Characterization of Birkhoff-James orthogonality

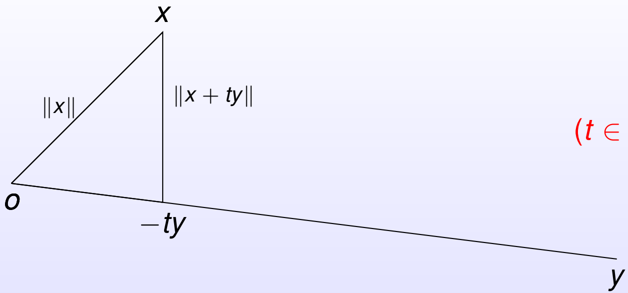
Priyanka Grover

December 15, 2014

Motivation

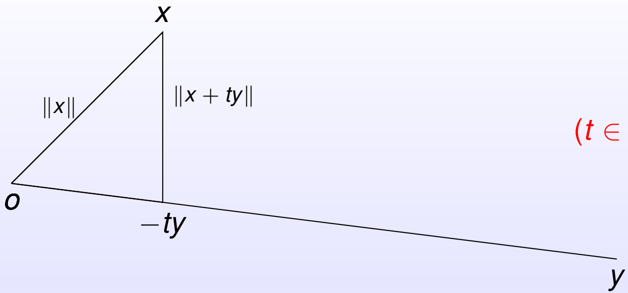


Motivation



$(t \in \mathbb{R})$

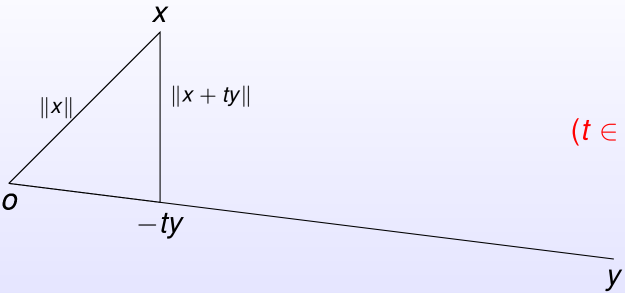
Motivation



$(t \in \mathbb{R})$

Either $\|x+ty\| \geq \|x\|$ for all $t \geq 0$ or $\|x+ty\| \geq \|x\|$ for all $t \leq 0$.

Motivation



$(t \in \mathbb{R})$

Either $\|x+ty\| \geq \|x\|$ for all $t \geq 0$ or $\|x+ty\| \geq \|x\|$ for all $t \leq 0$.

$\|x+ty\| \geq \|x\|$ for all $t \in \mathbb{R}$ if and only if x and y are orthogonal.

Birkhoff-James orthogonality

$\mathscr{X}$ complex Banach space
$x, y \in \mathscr{X}$

x is said to be Birkhoff-James orthogonal to y ($x \perp_{BJ} y$) if

$$\|x + \lambda y\| \geq \|x\| \text{ for all } \lambda \in \mathbb{C}.$$

- When $\mathscr{X}$ is a Hilbert space, this is the same as usual orthogonality.

Note that we can also have

$$\|x + ty\| \geq \|x\| \text{ for all } t \in \mathbb{R}.$$

Notation $x \perp_{BJ}^{(real)} y$

Properties

- This orthogonality is clearly homogeneous: x orthogonal to $y \Rightarrow \lambda x$ orthogonal to μy for all scalars λ, μ.
- Not symmetric: x orthogonal to $y \not\Rightarrow y$ orthogonal to x.
- Not additive: x orthogonal to $y, z \not\Rightarrow x$ orthogonal to $y + z$.

New method

$$\|x + ty\| \geq \|x\| \text{ for all } t \in \mathbb{R}$$

- Let $f(t) = \|x + ty\|$ mapping $\mathbb{R}$ into $\mathbb{R}_+$.
- To say that $\|x + ty\| \geq \|x\|$ for all $t \in \mathbb{R}$ is to say that f attains its minimum at the point 0.
- A calculus problem?
- If f were differentiable, then a necessary and sufficient condition for this would have been that the derivative $\mathrm{D}\, f(0) = 0$.
- But the norm function may not be differentiable at x.
- However, f is a convex function, that is,

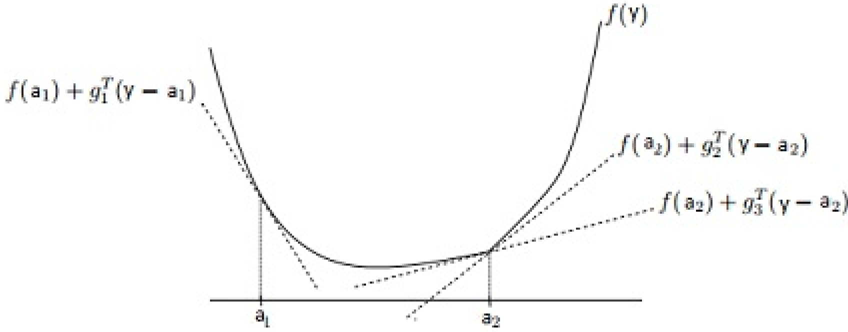
  $f(\alpha x + (1-\alpha) y) \leq \alpha\, f(x) + (1-\alpha)\, f(y)$ for all $x, y \in \mathscr{X}, 0 \leq \alpha \leq 1$.
- The tools of convex analysis are available.

Subdifferential

Definition

Let $f : \mathscr{X} \to \mathbb{R}$ be a convex function. The *subdifferential* of f at a point $a \in \mathscr{X}$, denoted by $\partial f(a)$, is the set of continuous linear functionals $\varphi \in \mathscr{X}^*$ such that

$$f(y) - f(a) \geq \operatorname{Re} \varphi(y - a) \quad \text{for all } y \in \mathscr{X}.$$

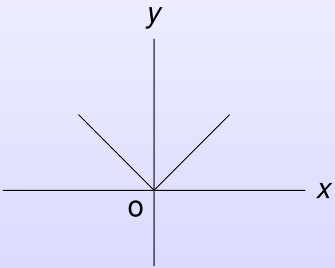


Examples

Let $f : \mathbb{R} \to \mathbb{R}$ be defined as

$$f(x) = |x|.$$

This function is differentiable at all $a \neq 0$ and $\mathrm{D}\, f(a) = sign(a)$. At zero, it is not differentiable.

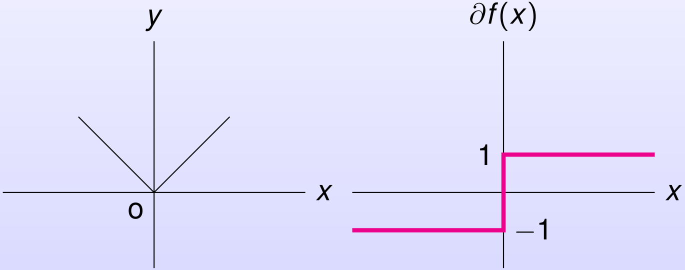


Examples

Let $f : \mathbb{R} \to \mathbb{R}$ be defined as

$$f(x) = |x|.$$

This function is differentiable at all $a \neq 0$ and $\mathrm{D}\, f(a) = sign(a)$. At zero, it is not differentiable.



$$\partial f(0) = [-1, 1].$$

Note that for $v \in \mathbb{R}$,

$$f(y) = |y| \geq f(0) + v.y = v.y$$

holds for all $y \in \mathbb{R}$ if and only if $|v| \leq 1$.

Examples

Let $f : \mathscr{X} \to \mathbb{R}$ be defined as

$$f(a) = \|a\|.$$

Then for $a \neq 0$,

$$\partial f(a) = \{\varphi \in \mathscr{X}^* : \text{ Re } \varphi(a) = \|a\|, \|\varphi\| \leq 1\},$$

and

$$\partial f(0) = \{\varphi \in \mathscr{X}^* : \|\varphi\| \leq 1\}.$$

Examples

Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined as

$$f(a) = \|a\|_\infty = \max\{|a_1|, \ldots, |a_n|\}.$$

Then for $a \neq 0$,

$$\partial f(a) = \text{ conv}\{\pm e_i : |a_i| = \|a\|_\infty\}.$$

$$\partial f(a) = \{\varphi \in \mathscr{X}^* : f(y) - f(a) \geq \text{ Re } \varphi(y - a) \quad \text{for all } y \in X\}.$$

Proposition

A convex function $f : \mathscr{X} \to \mathbb{R}$ attains its minimum value at $a \in \mathscr{X}$ if and only if $0 \in \partial f(a)$.

Positive combinations

Let $f_1, f_2 : \mathscr{X} \to \mathbb{R}$ be two convex functions and let t_1, t_2 be positive numbers. Then

$$\partial(t_1 f_1 + t_2 f_2)(a) = t_1 \partial f_1(a) + t_2 \partial f_2(a) \text{ for all } a \in \mathscr{X}.$$

Subdifferential calculus

Precomposition with an affine map

Let $\mathscr{X}, \mathscr{Y}$ be any two Banach spaces. Let $g : \mathscr{Y} \to \mathbb{R}$ be a convex function. Let $S : \mathscr{X} \to \mathscr{Y}$ be a linear map and let $L : \mathscr{X} \to \mathscr{Y}$ be the affine map defined by $L(x) = S(x) + y_0$, for some $y_0 \in \mathscr{Y}$. Then

$$\partial(g \circ L)(a) = S^* \partial g(L(a)) \text{ for all } a \in \mathscr{X}.$$

Birkhoff-James orthogonality by subdifferential calculus

$$\|x + \lambda y\| \geq \|x\| \text{ for all } \lambda \in \mathbb{C} \tag{1}$$

- Reduce the problem to solving $x \perp_{BJ}^{(real)} y$

(1) is equivalent to saying that for each fixed $\theta \in \mathbb{R}$

$$\|x + ty_\theta\| \geq \|x\| \text{ for all } t \in \mathbb{R},$$

where $y_\theta = e^{i\theta} y$

- Let $f(t) = \|x + ty\|$. Then $0 \in \partial f(0)$

$\|x + ty\| \geq \|x\|$ for all $t \in \mathbb{R} \Leftrightarrow f(t) \geq f(0)$ for all $t \in \mathbb{R} \Leftrightarrow$
$0 \in \partial f(0)$

Birkhoff-James orthogonality by subdifferential calculus

- f is precomposition with an affine map

$$S : t \mapsto ty$$

$$L : t \mapsto x + S(t) \quad \text{affine map}$$

$$g : a \mapsto \|a\| \quad \text{convex map}$$

$$f(t) = (g \circ L)(t)$$

- $0 \in S^* \partial \|x\|$

$$\partial f(0) = \partial (g \circ L)(0) = S^* \partial g(L(0)) = S^* \partial \|x\|$$

Birkhoff-James orthogonality by subdifferential calculus

$$\|x + ty\| \geq \|x\| \text{ for all } t \in \mathbb{R} \text{ if and only if } 0 \in S^*\partial\|x\|,$$

where $S(t) = ty$ for all $t \in \mathbb{R}$.

Orthogonality in matrices

$\mathbb{M}(n)$: the space of $n \times n$ complex matrices

$\langle A, B \rangle = \mathrm{tr}(A^* B)$

$\| \cdot \|$ is the operator norm, $\|A\| = \sup_{\|x\|=1} \|Ax\|$.

Theorem (Bhatia, Šemrl; 1999)

Let $A, B \in \mathbb{M}(n)$. Then $A \perp_{BJ} B$ if and only if there exists $x : \|x\| = 1,\ \|Ax\| = \|A\|$ and $\langle Ax, Bx \rangle = 0$.

Importance: It connects the more complicated Birkhoff-James orthogonality in the space $\mathbb{M}(n)$ to the standard orthogonality in the space $\mathbb{C}^n$.

Bhatia-Šemrl Theorem

Sufficient to prove that if $A \geq 0$, $\|A + tB\| \geq \|A\|$ for all $t \in \mathbb{R}$ if and only if there exists $x : \|x\| = 1,\ Ax = \|A\|x$ and Re $\langle Ax, Bx \rangle = 0$.

Let $A = U\Sigma V$ (U and V unitary matrices) be a singular value decomposition of A.

$$\|\Sigma + tU^*BV^*\| \geq \|\Sigma\| \text{ for all } t \in \mathbb{R}.$$

If there exists a unit vector y such that

$$\Sigma y = \|\Sigma\|y \text{ and } \operatorname{Re} \langle \Sigma y, U^*BV^*y \rangle = 0,$$

then for $x = V^*y$ we have

$$\|Ax\| = \|A\| \text{ and } \operatorname{Re} \langle Ax, Bx \rangle = 0.$$

Orthogonality in matrices

$A \geq 0$, $\|A + tB\| \geq \|A\|$ for all $t \in \mathbb{R}$ if and only if there exists $x : \|x\| = 1,\ Ax = \|A\|x$ and $\text{Re}\ \langle Ax, Bx \rangle = 0$.

- $S : \mathbb{R} \to \mathbb{M}(n)$ is the map $S(t) = tB$.
- $\|A + tB\| \geq \|A\|$ for all $t \in \mathbb{R}$ if and only if $0 \in S^* \partial \|A\|$, where $S^*(T) = \text{ Re } \text{tr}(B^* T)$

Watson, 1992

For $A \geq 0$

$$\partial \|A\| = \text{conv}\{uu^* : \|u\| = 1, Au = \|A\|u\}.$$

Bhatia-Šemrl Theorem

- $0 \in \partial f(0) = S^* \partial \|A\|$ if and only if $0 \in \text{conv}\{\text{ Re } \langle u, Bu \rangle : \|u\| = 1, Au = \|A\|u\}$.

- By Hausdorff-Toeplitz Theorem, $\{\text{ Re } \langle u, Bu \rangle : \|u\| = 1, Au = \|A\|u\}$ is convex.

- $0 \in S^* \partial \|A\|$ if and only if $0 \in \{\text{ Re } \langle u, Bu \rangle : \|u\| = 1, Au = \|A\|u\}$.

- There exists $x : \|x\| = 1,\ Ax = \|A\|x$ and $\text{ Re } \langle Ax, Bx \rangle = 0$.

Distance of A from $\mathbb{C}I$:

$$\mathrm{dist}(A, \mathbb{C}I) = \min\{\|A - \lambda I\| : \lambda \in \mathbb{C}\}$$

Variance of A with respect to x:
For $x : \|x\| = 1$,

$$\mathrm{var}_x(A) = \|Ax\|^2 - |\langle x, Ax\rangle|^2.$$

Corollary

Let $A \in \mathbb{M}(n)$. With notations as above, we have

$$\mathrm{dist}(A, \mathbb{C}I)^2 = \max_{\|x\|=1} \mathrm{var}_x(A).$$

Distance to $\mathbb{C}I$

Idea:

- Let $\operatorname{dist}(A, \mathbb{C}I) = \|A_0\|$, where $A_0 = A - \lambda_0 I$, for some $\lambda_0 \in \mathbb{C}$
- $A_0 \perp_{BJ} I$
- There exists $x : \|x\| = 1$ such that $\|A_0 x\| = \|A_0\|$ and $\langle x, A_0 x\rangle = 0$.
- $\operatorname{dist}(A, \mathbb{C}I)^2 = \|A_0\|^2 = \|A_0 x\|^2 = \|Ax\|^2 - |\langle x, Ax\rangle|^2$.
- $\operatorname{dist}(A, \mathbb{C}I)^2 \leq \max_{\|x\|=1} \operatorname{var}_x(A)$.
- For every $x : \|x\| = 1, \operatorname{var}_x(A) = \|Ax\|^2 - |\langle x, Ax\rangle|^2 \leq \|A\|^2$.
- Let $\lambda \in \mathbb{C}$. Change $A \to A - \lambda I$. Since variance is translation invariant, we get

$$\operatorname{var}_x(A) \leq \|A - \lambda I\|^2.$$

Orthogonality to a subspace

$\mathscr{W}$: subspace of $\mathbb{M}(n)$

A is said to be Birkhoff-James orthogonal to $\mathscr{W}$ ($A \perp_{BJ} \mathscr{W}$) if

$$\|A + W\| \geq \|A\| \text{ for all } W \in \mathscr{W}.$$

$\mathscr{W}^{\perp}$: the orthogonal complement of $\mathscr{W}$, under the usual Hilbert space orthogonality in $\mathbb{M}(n)$ with the inner product $\langle A, B\rangle = \operatorname{tr}(A^*B)$.

Bhatia-Šemrl theorem: $A \perp_{BJ} \mathbb{C}B$ if and only if there exists a positive semidefinite matrix P of rank one such that $\operatorname{tr} P = 1, \ \operatorname{tr} A^*AP = \|A\|^2$ and $AP \in (\mathbb{C}B)^{\perp}$.

$\mathbb{D}(n;\mathbb{R})$: the space of real diagonal $n \times n$ matrices

A matrix A is said to be minimal if $\|A + D\| \geq \|A\|$ for all $D \in \mathbb{D}(n;\mathbb{R})$, i.e. A is orthogonal to the subspace $\mathbb{D}(n;\mathbb{R})$.

Theorem (Andruchow, Larotonda, Recht, Varela; 2012)

A Hermitian matrix A is minimal if and only if there exists a $P \geq 0$ such that

$$A^2 P = \|A\|^2 P$$

and

all the diagonal elements of AP are zero.

Question: Similar characterizations for other subspaces?

Orthogonality to a subspace

Theorem

Let $A \in \mathbb{M}(n)$ *and let* $\mathscr{W}$ *be a subspace of* $\mathbb{M}(n)$. *Then* $A \perp_{BJ} \mathscr{W}$ *if and only if there exists* $P \geq 0$, $\operatorname{tr} P = 1$, *such that*

$$A^* AP = \|A\|^2 P$$

and

$$AP \in \mathscr{W}^{\perp}.$$

Moreover, we can choose P *such that* $\operatorname{rank} P \leq m(A)$, *where* $m(A)$ *is the multiplicity of the maximum singular value* $\|A\|$ *of* A.

Orthogonality to a subspace

$m(A)$ is the best possible upper bound on rank P.

Consider $\mathscr{W} = \{X : \operatorname{tr} X = 0\}$.

Then $\{A : A \perp_{BJ} \mathscr{W}\} = \mathscr{W}^{\perp} = \mathbb{C}I$.

If $A \perp_{BJ} \mathscr{W}$, then it has to be of the form $A = \lambda I$, for some $\lambda \in \mathbb{C}$.

When $A \neq 0$ then $m(A) = n$.

Let P be any density matrix satisfying $AP \in \mathscr{W}^{\perp}$. Then $AP = \mu I$, for some $\mu \in \mathbb{C}, \mu \neq 0$.

If P also satisfies $A^* AP = \|A\|^2 P$, then we get $P = \frac{\mu}{\lambda} I$. Hence rank $P = n = m(A)$.

Orthogonality to a subspace

Observation: In general, the set $\{A : A \perp_{BJ} \mathscr{W}\}$ need not be a subspace.

Consider the subspace $\mathscr{W} = \mathbb{C}I$ of $\mathbb{M}(3)$. Let

$$A_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ and } A_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Then $A_1, A_2 \perp_{BJ} \mathscr{W}$.

$$\text{Then } A_1 + A_2 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \|A_1 + A_2\| = 2.$$

But $\left\|A_1 + A_2 - \frac{1}{2}I\right\| = \frac{3}{2} < \|A_1 + A_2\|$. Hence $A_1 + A_2 \not\perp_{BJ} \mathscr{W}$.

Distance to any subalgebra of $\mathbb{M}(n)$

$\mathrm{dist}(A, \mathscr{W})$: distance of a matrix A from the subspace $\mathscr{W}$

$$\mathrm{dist}(A, \mathscr{W}) = \min\{\|A - W\| : W \in \mathscr{W}\}.$$

We have seen that

$$\mathrm{dist}(A, \mathbb{C}I)^2 = \max_{\|x\|=1} \mathrm{var}_x(A).$$

This is equivalent to saying that

$$\mathrm{dist}(A, \mathbb{C}I)^2 = \max\left\{\mathrm{tr}(A^*AP) - |\,\mathrm{tr}(AP)|^2 : P \geq 0, \mathrm{tr}\,P = 1, \mathit{rank}\ P = 1\right\}.$$

Let $\mathscr{B}$ be any C^*-subalgebra of $\mathbb{M}(n)$.

Similar distance formula?
(This question has been raised by Rieffel)

Distance to a subalgebra of $\mathbb{M}(n)$

$\mathcal{C}_{\mathscr{B}} : \mathbb{M}(n) \to \mathscr{B}$ denote the projection of $\mathbb{M}(n)$ onto $\mathscr{B}$.

Theorem

For any $A \in \mathbb{M}(n)$

$$\operatorname{dist}(A, \mathscr{B})^2 = \max\{\operatorname{tr}(A^*AP - \mathcal{C}_{\mathscr{B}}(AP)^*\mathcal{C}_{\mathscr{B}}(AP)\mathcal{C}_{\mathscr{B}}(P)^{-1}) \\ : P \geq 0, \operatorname{tr} P = 1\},$$

where $\mathcal{C}_{\mathscr{B}}(P)^{-1}$ denotes the Moore-Penrose inverse of $\mathcal{C}_{\mathscr{B}}(P)$. The maximum on the right hand side can be restricted to rank $P \leq m(A)$.

Orthogonality in C^*- algebra

Bhatia and Šemrl, 1999

$A, B \in \mathcal{B}(\mathcal{H})$, $A \perp_{BJ} B$ if and only if there exists a sequence $\{x_n\}$ of unit vectors such that $\|Ax_n\| \to \|A\|$, and $\langle Ax_n, Bx_n \rangle \to 0$ as $n \to \infty$.

Theorem

Let $\mathcal{A}$ be a C^*-algebra. Let $a, b \in \mathcal{A}$. Then $a \perp_{BJ} b$ if and only if there exists a state φ on $\mathcal{A}$ such that

$$\varphi(a^*a) = \|a\|^2 \text{ and } \varphi(a^*b) = 0.$$

Distance to $\mathbb{C}1$

$S(\mathcal{A})$: the state space of $\mathcal{A}$
$\varphi \in S(\mathcal{A})$.

Let *variance* of a with respect to φ, denoted by $\mathrm{var}_\varphi(a)$, be defined as

$$\mathrm{var}_\varphi(a) = \varphi(a^* a) - |\varphi(a)|^2.$$

Theorem (Rieffel, 2012)

Let $a \in \mathcal{A}$. Let $S(\mathcal{A})$ denote the state space of $\mathcal{A}$.

$$\mathrm{dist}(a, \mathbb{C}1)^2 = \max\{\mathrm{var}_\varphi(a) : \varphi \in S(\mathcal{A})\}.$$

When $\mathcal{A} = \mathbb{M}(n)$, then

$$\mathrm{dist}(A, \mathbb{C}I)^2 = \max\left\{\mathrm{tr}(A^* AP) - |\,\mathrm{tr}(AP)|^2 : P \geq 0, \mathrm{tr}\, P = 1\right\}.$$

Hilbert C^*-modules

Let $\mathcal{A}$ be a C*-algebra. An inner-product $\mathcal{A}$-module is a vector space $\mathscr{E}$ which is a (right) $\mathcal{A}$-module (with compatible scalar multiplication:

$$\lambda(xa) = (\lambda x)a = x(\lambda a) \text{ for } x \in \mathscr{E}, a \in \mathcal{A}, \lambda \in \mathbb{C}),$$

together with a map $(x, y) \mapsto \langle x, y \rangle : \mathscr{E} \times \mathscr{E} \to \mathcal{A}$ such that

(i) $\langle x, \alpha y + \beta z \rangle = \alpha \langle x, y \rangle + \beta \langle x, z \rangle$ for all $x, y, z \in \mathscr{E}$, $\alpha, \beta \in \mathbb{C}$
(ii) $\langle x, ya \rangle = \langle x, y \rangle a$ for all $x, y \in \mathscr{E}$, $a \in \mathcal{A}$
(iii) $\langle y, x \rangle = \langle x, y \rangle^*$ for all $x, y \in \mathscr{E}$
(iv) $\langle x, x \rangle \geq 0$; if $\langle x, x \rangle = 0$ then $x = 0$.

For $x \in \mathscr{E}$,

$$\|x\| = \|\langle x, x \rangle\|^{1/2}.$$

An inner-product $\mathcal{A}$-module which is complete with respect to this norm is called a Hilbert $\mathcal{A}$-module.

Example: $\mathbb{M}(m,n)$ is a Hilbert $\mathbb{M}(n)$-module, with the inner product

$$\langle A, B\rangle = A^*B \text{ for all } A, B \in \mathbb{M}(m,n).$$

Similarly for infinite dimensional Hilbert spaces $\mathcal{H}, \mathcal{K}$,

$\mathcal{B}(\mathcal{H}, \mathcal{K})$ is a Hilbert $\mathcal{B}(\mathcal{H})$-module.

Theorem

Let $\mathcal{H}$ and $\mathcal{K}$ be two Hilbert spaces. Let $A, B \in \mathcal{B}(\mathcal{H}, \mathcal{K})$. Then $A \perp_{BJ} B$ if and only if there exists a state φ on $\mathcal{B}(\mathcal{H})$ such that $\varphi(A^*A) = \|A\|^2$ and $\varphi(A^*B) = 0$.

Let $\mathscr{E}$ be a (right) Hilbert $\mathcal{A}$-module.

Theorem (Blecher; 1997)

$\mathscr{E}$ can be isometrically embedded in $\mathscr{B}(\mathcal{H}, \mathcal{K})$ for some Hilbert spaces $\mathcal{H}, \mathcal{K}$.

As a consequence, we obtain the following.

Theorem

Let $e_1, e_2 \in \mathscr{E}$. Then $e_1 \perp_{BJ} e_2$ if and only if there exists a state φ on $\mathcal{A}$ such that

$$\varphi(\langle e_1, e_1\rangle) = \|e_1\|^2 \text{ and } \varphi(\langle e_1, e_2\rangle) = 0.$$

REFERENCES:

1. T. Bhattacharyya, P. Grover, Characterization of Birkhoff-James orthogonality, *J. Math. Anal. Appl.* 407 (2013) 350–358.
2. P. Grover, Orthogonality to matrix subspaces, and a distance formula, *Linear Algebra Appl.* 445 (2014) 280–288.
3. P. Grover, *Some problems in differential and subdifferential calculus of matrices*, Ph. D. thesis, Indian Statistical Institute, 2014.

THANK YOU