Orthogonality to matrix subspaces

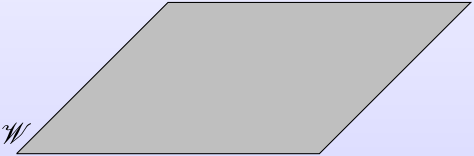
Priyanka Grover

Indian Institute of Technology, Delhi

April 2, 2015

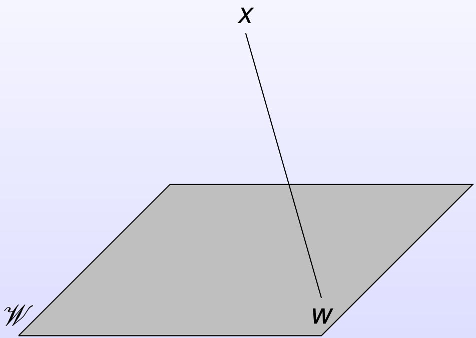
Motivation

Let $\mathscr{W}$ be a subspace of $\mathbb{R}^n$.



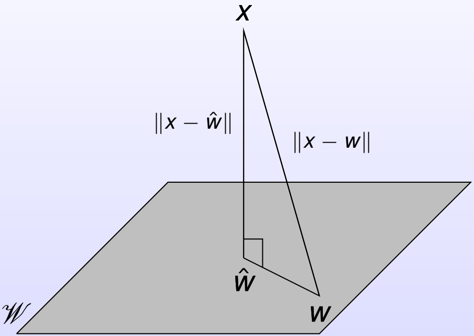
Motivation

Let $\mathscr{W}$ be a subspace of $\mathbb{R}^n$. Let $x \in \mathbb{R}^n$ be any point.



Motivation

Let $\mathscr{W}$ be a subspace of $\mathbb{R}^n$. Let $x \in \mathbb{R}^n$ be any point.

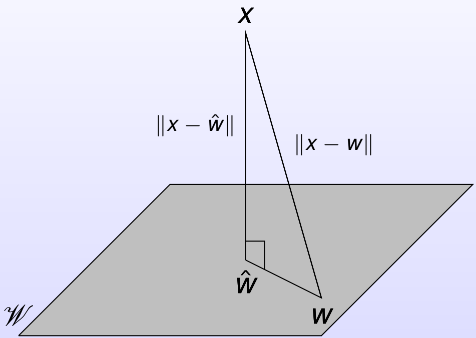


The orthogonal projection of x onto $\mathscr{W}$, namely $\hat{w}$ is the closest point in $\mathscr{W}$ to x.

$$\|x - \hat{w}\| < \|x - w\| \text{for all } w \in \mathscr{W}, w \neq \hat{w}$$

Motivation

Let $\mathscr{W}$ be a subspace of $\mathbb{R}^n$. Let $x \in \mathbb{R}^n$ be any point.



The orthogonal projection of x onto $\mathscr{W}$, namely $\hat{w}$ is the closest point in $\mathscr{W}$ to x.

$$\|x - \hat{w}\| < \|x - w\| \text{for all } w \in \mathscr{W}, w \neq \hat{w}$$

$\hat{w}$ is called the best approximation from $\mathscr{W}$ to x.

Motivation

Let $\mathbb{M}(n)$ be the space of $n \times n$ complex matrices. Let $\mathscr{W}$ be a subspace of $\mathbb{M}(n)$. Let $A \notin \mathscr{W}$. Consider the problem of finding a best approximation from $\mathscr{W}$ to A.

That is, find an element $\hat{W}$ such that

$$\min_{W \in \mathscr{W}} \|A - W\| = \|A - \hat{W}\|.$$

A specific question: When is zero a best approximation from $\mathscr{W}$ to A? That is, when do we have

$$\min_{W \in \mathscr{W}} \|A - W\| = \|A\|?$$

Suppose $\mathscr{W}$ is the subspace spanned by a single element B. Then the above problem reduces to

$$\min_{\lambda \in \mathbb{C}} \|A - \lambda B\| = \|A\|.$$

Motivation

Let $\mathbb{M}(n)$ be the space of $n \times n$ complex matrices. Let $\mathscr{W}$ be a subspace of $\mathbb{M}(n)$. Let $A \notin \mathscr{W}$. Consider the problem of finding a best approximation from $\mathscr{W}$ to A.
That is, find an element $\hat{W}$ such that

$$\min_{W \in \mathscr{W}} \|A - W\| = \|A - \hat{W}\|.$$

A specific question: When is zero a best approximation from $\mathscr{W}$ to A? That is, when do we have

$$\min_{W \in \mathscr{W}} \|A - W\| = \|A\|?$$

Suppose $\mathscr{W}$ is the subspace spanned by a single element B. Then the above problem reduces to

$$\min_{\lambda \in \mathbb{C}} \|A - \lambda B\| = \|A\|.$$

Motivation

Let $\mathbb{M}(n)$ be the space of $n \times n$ complex matrices. Let $\mathscr{W}$ be a subspace of $\mathbb{M}(n)$. Let $A \notin \mathscr{W}$. Consider the problem of finding a best approximation from $\mathscr{W}$ to A.
That is, find an element $\hat{W}$ such that

$$\min_{W \in \mathscr{W}} \|A - W\| = \|A - \hat{W}\|.$$

A specific question: When is zero a best approximation from $\mathscr{W}$ to A? That is, when do we have

$$\min_{W \in \mathscr{W}} \|A - W\| = \|A\|?$$

Suppose $\mathscr{W}$ is the subspace spanned by a single element B. Then the above problem reduces to

$$\min_{\lambda \in \mathbb{C}} \|A - \lambda B\| = \|A\|.$$

Motivation

Let $\mathbb{M}(n)$ be the space of $n \times n$ complex matrices. Let $\mathscr{W}$ be a subspace of $\mathbb{M}(n)$. Let $A \notin \mathscr{W}$. Consider the problem of finding a best approximation from $\mathscr{W}$ to A.
That is, find an element $\hat{W}$ such that

$$\min_{W \in \mathscr{W}} \|A - W\| = \|A - \hat{W}\|.$$

A specific question: When is zero a best approximation from $\mathscr{W}$ to A? That is, when do we have

$$\min_{W \in \mathscr{W}} \|A - W\| = \|A\|?$$

Suppose $\mathscr{W}$ is the subspace spanned by a single element B. Then the above problem reduces to

$$\min_{\lambda \in \mathbb{C}} \|A - \lambda B\| = \|A\|.$$

Definition

A is said to be Birkhoff-James orthogonal to B if

$$\min_{\lambda \in \mathbb{C}} \|A - \lambda B\| = \|A\|,$$

that is,

$$\|A + \lambda B\| \geq \|A\| \text{ for all } \lambda \in \mathbb{C}.$$

In general, if $\mathscr{X}$ is a complex normed linear space, then an element x is said to be Birkhoff-James orthogonal to another element y if

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

If $\mathscr{X}$ is a pre-Hilbert space with inner product $\langle \cdot, \cdot \rangle$, then this definition is equivalent to $\langle x, y \rangle = 0$.

Definition

A is said to be Birkhoff-James orthogonal to B if

$$\min_{\lambda \in \mathbb{C}} \|A - \lambda B\| = \|A\|,$$

that is,

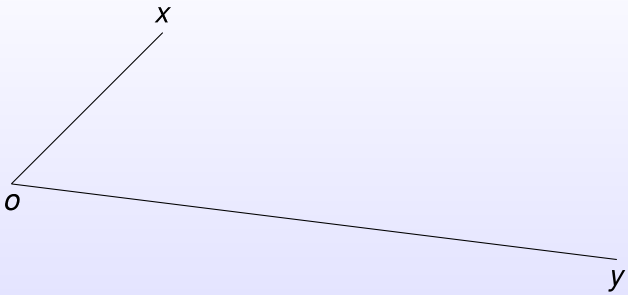
$$\|A + \lambda B\| \geq \|A\| \text{ for all } \lambda \in \mathbb{C}.$$

In general, if $\mathscr{X}$ is a complex normed linear space, then an element x is said to be Birkhoff-James orthogonal to another element y if

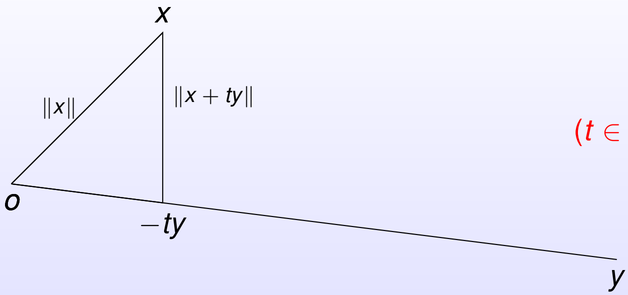
$$\|x + \lambda y\| \geq \|x\| \text{ for all } \lambda \in \mathbb{C}.$$

If $\mathscr{X}$ is a pre-Hilbert space with inner product $\langle \cdot, \cdot \rangle$, then this definition is equivalent to $\langle x, y \rangle = 0$.

Geometric interpretation

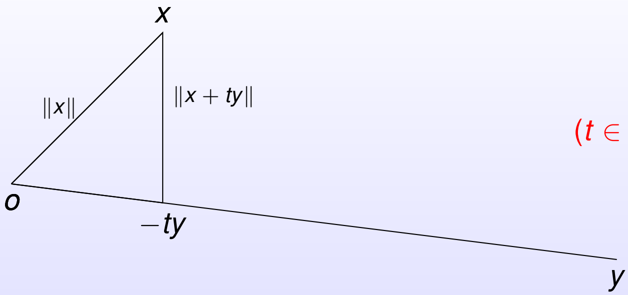


Geometric interpretation



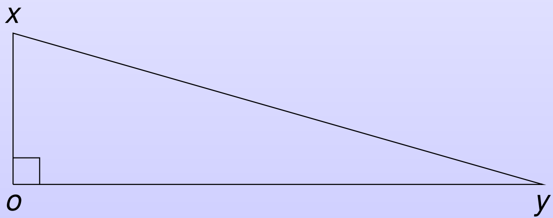
$(t \in \mathbb{R})$

Geometric interpretation



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

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



Birkhoff-James orthogonality

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

Observation

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

x is Birkhoff-James orthogonal to y if both of them are satisfied.

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 $y \nRightarrow y$ orthogonal to x.
- Not additive: x orthogonal to $y, z \nRightarrow x$ 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 at x, 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.

Orthogonality in matrices

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

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

$\| \cdot \|$ 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

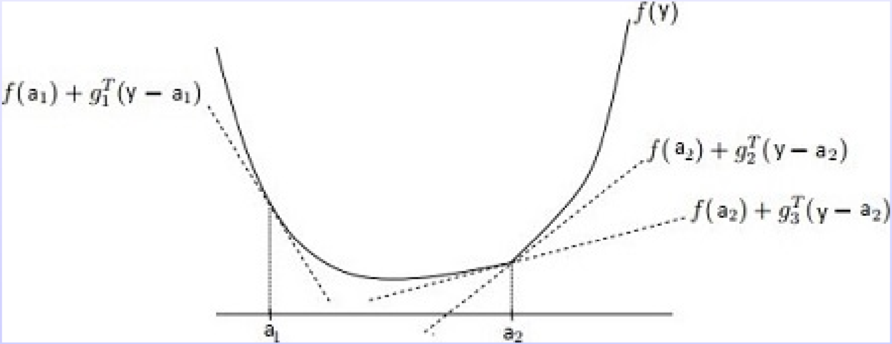
- Let $f : t \to \|A + tB\|$, $A \geq 0$.
- $A \perp_{BJ} B$ if and only if f attaines its minimum at 0.
- f is a convex function but may not be differentiable.

Subdifferential

Definition 1

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

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

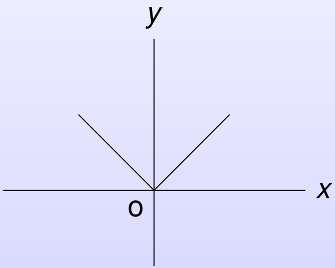


Examples

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.

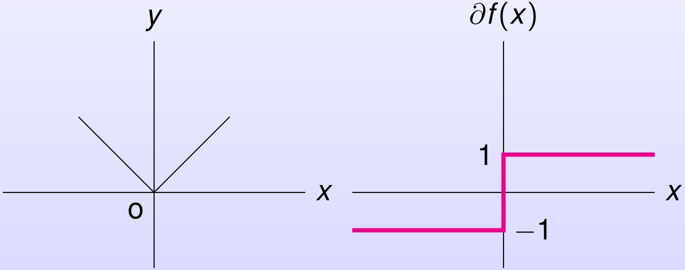


Examples

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 $v \in \mathbb{R}$,

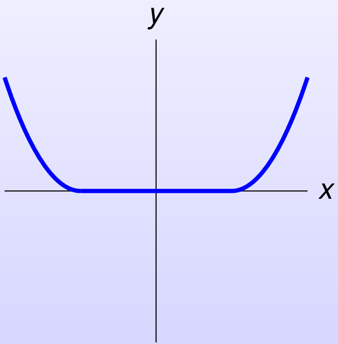
$$f(y) = |y| \geq f(0) + v.y = v.y$$

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

Examples

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

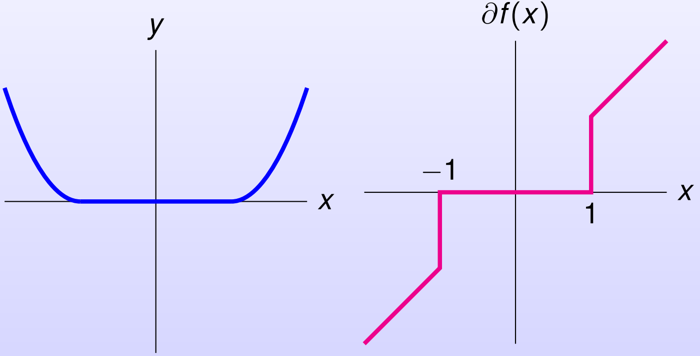
$$f(x) = \max\left\{0, \frac{x^2 - 1}{2}\right\}.$$



Examples

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

$$f(x) = \max\left\{0, \frac{x^2 - 1}{2}\right\}.$$



Then f is differentiable everywhere except at $x = -1, 1$. We have

$$\partial f(1) = [0, 1] \text{ and } \partial f(-1) = [-1, 0].$$

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

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

In particular, when $\mathscr{X} = \mathbb{M}(n)$, we get that for $A \neq 0$,

$$\partial f(A) = \{G \in \mathbb{M}(n) : \text{ Re } \operatorname{tr} G^* A = \|A\|, \|G\|_* \leq 1\}.$$

($\|\cdot\|_*$ is the dual norm of $\|\cdot\|$)

Examples

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

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

Then for $a \neq 0$,

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

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

Proposition

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

In our case,

$$f(t) = \|A + tB\|$$

and f attains minimum at 0. So

$$0 \in \partial f(0).$$

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

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

In our case, let $S : t \mapsto tB, \qquad L : t \mapsto A + tB$ and

$$g : X \mapsto \|X\|.$$

And $f(t) = \|A + tB\|$ is the composition of g and L. So

$$\partial f(0) = S^* \partial \|A\|,$$

where $S^*(T) = \text{ Re } \operatorname{tr} B^* T$.

Bhatia-Šemrl Theorem

Watson, 1992

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

$$\partial \|A\| = \operatorname{conv}\{uv^* : \|u\| = \|v\| = 1, Av = \|A\|u\}.$$

If $A \geq 0$, then

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

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 $\text{ Re } \langle Ax, Bx\rangle = 0$.

Application

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

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

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

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

Corollary

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

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

Orthogonality to a subspace

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

Consider the problem of finding $\hat{W}$ such that

$$\min_{W \in \mathscr{W}} \|A - W\| = \|A - \hat{W}\|$$

Examples

1) Let $\mathscr{W} = \{W : \operatorname{tr} W = 0\}$.

Then

$$\min_{W \in \mathscr{W}} \|A - W\| = \frac{|\operatorname{tr}(A)|}{n}.$$

$$\hat{W} = A - \frac{\operatorname{tr}(A)}{n} I.$$

2) Let $\mathscr{W} = \{W : W = W^*\}$.

For any A,

$$\hat{W} = \frac{1}{2}(A + A^*) = \operatorname{Re} A.$$

Orthogonality to a subspace

When is zero matrix a best approximation to $\mathscr{W}$?

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

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 $P \leq m(A)$.

The quantity on the right also enjoys the property of being translation invariant.

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