

# A GENERALIZED KARAMATA INEQUALITY

ETHAN Y. JAFFE

## 1. KARAMATA'S INEQUALITY

In this note we prove a generalized version of Karamata's inequality, and give an example of an application of the classical inequality to functional analysis.

First, some notation.

If  $z \in \mathbf{R}^n$ , write  $z^\downarrow \in \mathbf{R}^n$  for the vector which has the same components as  $z$ , but written in non-increasing order. For  $x, y \in \mathbf{R}^n$ , we say that “ $y$  weakly majorizes  $x$ ” if for all  $1 \leq k \leq n$

$$\sum_{i=1}^k x_i^\downarrow \leq \sum_{i=1}^k y_i^\downarrow,$$

and write  $x \preceq y$ . If in addition  $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$ , then we say “ $y$  strongly majorizes  $x$ ” and write  $x \prec y$ . Observe that the condition  $x \preceq y$  is equivalent to the following condition: for any  $1 \leq \ell \leq n$ , if  $i_1, \dots, i_\ell, j_1, \dots, j_\ell$  are indices of the  $\ell$  largest elements of  $x$  and  $y$ , respectively, then

$$\sum_{p=1}^{\ell} x_{i_p} \leq \sum_{p=1}^{\ell} y_{j_p}.$$

For  $y \in \mathbf{R}^n$ , denote  $C(y) = \{z \in \mathbf{R}^n : z^\downarrow \prec y^\downarrow\}$ .

Let the permutation group  $S_n$  act on  $\mathbf{R}^n$  via

$$\sigma \cdot x = x^\sigma = (x_{\sigma(1)}, \dots, x_{\sigma(n)}),$$

for  $\sigma \in S_n$  and  $x \in \mathbf{R}^n$ . If  $A \subseteq \mathbf{R}^n$ , let  $\text{Con}(A)$  denote its convex hull. If  $x, y \in \mathbf{R}^n$  then we write  $x \leq y$  if  $x_i \leq y_i$  for all  $i$ .

**Theorem 1.1** (Generalized Karamata's inequality). *Suppose  $K \subseteq \mathbf{R}^n$  is convex and invariant under the action of  $S_n$ . Let  $\Phi : K \rightarrow \mathbf{R}$  be convex and symmetric. Fix  $x, y \in K$ . Then, if  $x \prec y$ ,*

$$\Phi(x) \leq \Phi(y).$$

*If we assume further that  $\Phi$  is non-decreasing in the sense that  $u \leq v$  implies  $\Phi(u) \leq \Phi(v)$ , then if  $x \preceq y$ ,*

$$\Phi(x) \leq \Phi(y).$$

.

*Remark 1.2.* The classical Karamata's inequality is the special case that  $C = \mathbf{R}^n$  and  $\Phi(z) := \sum_{i=1}^n \varphi(z_i)$  for some (nondecreasing) convex function  $\varphi$  of one real variable.

The proof will examine the convexity properties of the set  $C(y)$ . The theorem will follow quickly from the following proposition:

**Proposition 1.3.** For  $y \in \mathbf{R}^n$ ,  $C(y)$  is convex and compact, and invariant under the action of  $S_n$ . The set of extreme points<sup>1</sup> of  $C(y)$  is precisely the set

$$\{y^\sigma : \sigma \in S_n\}.$$

We now prove the generalized Karamata's inequality, given proposition 1.3.

*Proof of theorem 1.1.* Since  $y \in K$  and  $K$  is invariant under the  $S_n$  action,

$$\{y^\sigma : \sigma \in S_n\} \subseteq K,$$

and thus by convexity

$$\text{Con}(\{y^\sigma : \sigma \in S_n\}) \subseteq K.$$

By the Krein-Milman theorem<sup>2</sup> and proposition 1.3,  $C(y)$  is the convex hull of its extreme points, i.e.

$$C(y) = \text{Con}(\{y^\sigma : \sigma \in S_n\}) \subseteq K.$$

Thus  $C(y) \cap K = C(y)$ . Assume first that  $x \prec y$ . Then

$$x \in C(y) \cap K = C(y) = \text{Con}(\{y^\sigma : \sigma \in S_n\}),$$

and so  $x$  is a convex combination of the points  $y^\sigma$ , i.e. for each  $\sigma \in S_n$  there exists  $0 \leq t_\sigma \leq 1$  with  $\sum_{\sigma \in S_n} t_\sigma = 1$  and

$$x = \sum_{\sigma \in S_n} t_\sigma y^\sigma.$$

Since  $\Phi$  is convex and symmetric,

$$\Phi(x) \leq \sum_{\sigma \in S_n} t_\sigma \Phi(y^\sigma) = \Phi(y),$$

which is what we wanted to prove.

Assume now that  $x \preceq y$  and that  $\Phi$  is non-decreasing. Since  $\Phi$  is symmetric, we may suppose that  $x = x^\downarrow$  and  $y = y^\downarrow$ . The proof will be complete if we can show:

**Lemma.** *There exists  $z \in C(y) = C(y) \cap K$  such that  $x \leq z$ .*

For if this is true, then  $\Phi(x) \leq \Phi(z)$  since  $\Phi$  is non-decreasing, and  $\Phi(z) \leq \Phi(y)$  by the first part of the theorem since  $z \prec y$  by definition.

*Proof of lemma.* We use induction on  $n$ . If  $n = 1$ , then  $x \preceq y$  is equivalent to  $x \leq y$ , so the lemma holds trivially. We prove the lemma for  $n$ , assuming it holds true for  $n - 1$ . Write  $x', y' \in \mathbf{R}^{n-1}$  for the vectors obtained by deleting the last coordinate of  $x, y$ , respectively. By assumption,  $x'^\downarrow = x'$ ,  $y'^\downarrow = y'$ , and  $x' \preceq y'$ . By induction there is

$$z' \in C(y')$$

such that  $x' \leq z'$ . As we showed above,

$$C(y') = \text{Con}(\{y'^\tau : \tau \in S_{n-1}\}),$$

<sup>1</sup>Recall that an extreme point  $e$  of a convex set  $K$  is one which lies on no proper line segment in  $K$ , i.e. if  $e = (1 - t)p + tq$  for  $p, q \in K$  and  $t \in (0, 1)$ , then  $p = q = e$ .

<sup>2</sup>See theorem A.1 for proof. Notice that we are working in finite dimensions, and so we do not need to take the closure of the convex hull, although this would not really affect the proof, anyway.

and so there is for each  $\tau \in S_{n-1}$ , a  $0 \leq t_\tau \leq 1$  such that  $\sum_{\tau \in S_{n-1}} t_\tau = 1$  and

$$x' \leq z' = \sum_{\tau \in S_{n-1}} t_\tau y'^\tau.$$

Extend  $\tau$  to  $S_n$  by letting by setting  $\tau(n) = n$ , and define  $z := \sum_{\tau \in S_{n-1}} t_\tau y^\tau \in C(y)$ . Then  $z_n = y_n$ . If  $x_n \leq y_n$ , then it follows from  $x' \leq z'$  that  $x \leq z$ , and we are done. Otherwise,  $x_n > y_n$ . Let us consider

$$t = \frac{z_1 + \cdots + z_n - (x_1 + \cdots + x_n)}{z_1 + \cdots + z_{n-1} - (x_1 + \cdots + x_{n-1})}.$$

Notice that the denominator is nonzero. Indeed, if it were not, then since  $z_i - x_i \geq 0$  for  $1 \leq i \leq n-1$  by assumption,  $z_i - x_i = 0$  for all  $1 \leq i \leq n-1$ , and hence

$$x_1 + \cdots + x_{n-1} + x_n = z_1 + \cdots + z_{n-1} + x_n = y_1 + \cdots + y_{n-1} + x_n > y_1 + \cdots + y_{n-1} + y_n,$$

which contradicts  $x \preccurlyeq y$ . The numerator is non-negative, since

$$x_1 + \cdots + x_n \leq y_1 + \cdots + y_{n-1} + y_n = z_1 + \cdots + z_{n-1} + z_n.$$

However, since  $z_n - x_n < 0$  by assumption, the numerator is at most the denominator. We deduce that  $0 \leq t \leq 1$ . Consider  $w \in \mathbf{R}^{n-1}$  defined by

$$w_i = \begin{cases} x_i + (z_i - x_i)t, & 1 \leq i \leq n-1 \\ x_n, & i = n. \end{cases}$$

Since  $0 \leq t \leq 1$ ,  $x \leq w$ . Furthermore, for  $1 \leq \ell \leq n-1$ , suppose  $i_1, \dots, i_\ell$  are the indices of the  $\ell$ -largest elements of  $w_i$ . Then we may assume that no  $i_j$  is  $n$ , since  $w_n = x_n \leq x_i \leq w_i$  for all  $i < n$ . Thus, since  $0 \leq t \leq 1$ ,

$$\sum_{j=1}^{\ell} w_{i_j} = \sum_{j=1}^{\ell} x_{i_j} + t(z_{i_j} - x_{i_j}) \leq \sum_{j=1}^{\ell} z_{i_j} \leq \sum_{j=1}^{\ell} z_j^\downarrow \leq \sum_{j=1}^{\ell} y_j,$$

with the last inequality holding since  $z' \in C(y')$ . But by definition of  $t$ ,

$$\sum_{j=1}^n w_j = (x_1 + \cdots + x_n) + t(z_1 + \cdots + z_{n-1} - (x_1 + \cdots + x_{n-1})) = z_1 + \cdots + z_n = y_1 + \cdots + y_n.$$

These last two statements taken together mean that  $w \in C(y)$ . Thus we have found  $x \leq w \in C(y)$ . □

□

*Proof of proposition 1.3.* Write  $C = C(y)$ . Without loss of generality, we may assume that  $y = y^\downarrow$ . Then  $C$  is by definition invariant under the  $S_n$  action. To show convexity, suppose  $u, v \in C$ . Fix  $0 \leq t \leq 1$ , and suppose  $i_1, \dots, i_k$  are the indices of the  $k$ -largest elements of  $z = (1-t)u + tv$ . Then

$$\sum_{j=1}^k z_{i_j} \leq (1-t) \sum_{j=1}^k u_{i_j} + t \sum_{j=1}^k v_{i_j} \leq (1-t) \sum_{j=1}^k u_j^\downarrow + t \sum_{j=1}^k v_j^\downarrow \leq \sum_{j=1}^k y_j.$$

By the same reasoning,  $\sum_{j=1}^n z_j = \sum_{j=1}^n y_j$ . Thus  $z \in C$ .

The set  $C$  is certainly closed, so to show compactness, it suffices to find uniform bounds of the form  $a \leq z_i \leq A$  for all  $z \in C$  and  $1 \leq i \leq n$ . Fix  $z \in C$ , and without loss of generality assume  $z = z^\downarrow$  (the uniform bounds are invariant under permutation). Certainly  $z_i \leq z_1 \leq y_1$  for all  $1 \leq i \leq n$ , so  $y_1$  is an upper bound. However,  $z_i \geq z_n = \sum_{i=1}^{n-1} y_i - z_i + y_n \geq y_n$ , for all  $1 \leq i \leq n$ , so  $y_n$  is a lower bound.

Let us now show that each  $y^\sigma$  for  $\sigma \in S_n$  is an extreme point of  $C$ . Since  $C$  is invariant under the  $S_n$  action, we may assume that  $\sigma$  is the identity. Suppose  $y = (1-t)x + tz$  for  $x, z \in C$ , and  $0 < t < 1$ . Then, for all  $1 \leq i \leq n$ ,

$$\sum_{i=1}^k y_i = (1-t) \sum_{i=1}^k x_i + t \sum_{i=1}^k z_i \leq (1-t) \sum_{i=1}^k x_i^\downarrow + t \sum_{i=1}^k z_i^\downarrow \leq \sum_{i=1}^k y_i.$$

Thus equality holds throughout, and  $\sum_{i=1}^k x_i = \sum_{i=1}^k x_i^\downarrow$ ,  $\sum_{i=1}^k z_i = \sum_{i=1}^k z_i^\downarrow$  and  $\sum_{i=1}^k x_i^\downarrow = \sum_{i=1}^k z_i^\downarrow = \sum_{i=1}^k y_i$ . Thus  $\sum_{i=1}^k x_i = \sum_{i=1}^k z_i = \sum_{i=1}^k y_i$  for all  $1 \leq k \leq n$ . Starting with  $k = 1$ , which yields  $x_1 = z_1 = y_1$ , one proceeds inductively to show that  $x_i = z_i = y_i$  for all  $1 \leq i \leq n$ .

We will prove that these are the only extreme points by induction on  $n$ . If  $n = 1$ , then  $C(y) = \{y\}$ , and the only extreme point is just  $y$  itself. To perform the inductive step, we will need a lemma on breaking up and reassembling the set  $C(y)$ . Fix  $1 \leq k \leq n$ , and write any  $z \in \mathbf{R}^n$  as  $(z', z'') \in \mathbf{R}^k \times \mathbf{R}^{n-k}$ . Define  $C' = C(y_1, \dots, y_k)$  and  $C'' = C(y_{k+1}, \dots, y_n)$  (if  $k = n$ , then leave  $C'' := \emptyset$ ).

**Lemma.** *If  $z' \in C'$  and  $z'' \in C''$ , then  $z = (z', z'') \in C$ . Conversely, if  $z = (z', z'') \in C$ , then  $z' \in C'$ .*

*Suppose additionally that  $\sum_{i=1}^k z_i = \sum_{i=1}^k y_i$ . Then  $z'' \in C''$ . Moreover, if  $z \in C$  is extreme, then  $z' \in C'$  and  $z'' \in C''$  are extreme.*

*Proof.* Suppose  $z' \in C'$  and  $z'' \in C''$ , and set  $z = (z', z'')$ . Let  $1 \leq \ell \leq k$  and suppose  $i_1, \dots, i_\ell$  are the indices of the  $\ell$  largest elements of  $z$ . Suppose  $p$  of the indices are in  $\{1, \dots, k\}$  and  $q = \ell - p$  of the are in  $\{k+1, \dots, n\}$ . Then

$$\begin{aligned} \sum_{j=1}^{\ell} z_{i_j} &= \sum_{i_\ell \in \{1, \dots, k\}} z_{i_\ell} + \sum_{i_\ell \in \{k+1, \dots, n\}} z_{i_\ell} \\ &= \sum_{i_\ell \in \{1, \dots, k\}} z'_{i_\ell} + \sum_{i_\ell \in \{k+1, \dots, n\}} z''_{i_\ell - p} \\ &\leq \sum_{i=1}^p z'_i^\downarrow + \sum_{j=1}^q z''_j^\downarrow \\ &\leq \sum_{i=1}^p y_i + \sum_{j=k+1}^{k+q} y_j \leq \sum_{i=1}^{p+q=\ell} y_i. \end{aligned}$$

Similarly,

$$\sum_{j=1}^n z_j = \sum_{j=1}^p z'_j + \sum_{j=1}^q z''_j = \sum_{j=1}^p y_j + \sum_{j=k+1}^{k+1} y_j = \sum_{j=1}^{p+q=\ell} y_j.$$

Thus  $z \in C$ .

Now suppose  $z \in C$ , and  $\sum_{i=1}^k z_i = \sum_{i=1}^k y_i$ . Split  $z = (z', z'') \in \mathbf{R}^k \times \mathbf{R}^{n-k}$ . Let us start by showing that  $z' \in C'$ . Indeed, if  $1 \leq \ell \leq k$  and  $i_1, \dots, i_\ell$  are the indices of the  $\ell$  largest elements of  $z'$ , then

$$\sum_{i=1}^{\ell} z'_i = \sum_{i=1}^{\ell} z_i \leq \sum_{i=1}^{\ell} z_i^{\downarrow} \leq \sum_{i=1}^{\ell} y_i.$$

If  $\ell = k$ , then equality must hold by assumption. To show that  $z'' \in C''$ , we need to show  $\sum_{j=1}^{n-k} z''_j = \sum_{j=1}^{n-k} y''_j$  and if  $1 \leq \ell \leq n-k$  and  $i_1, \dots, i_\ell$  are the indices of the  $\ell$  largest elements of  $z''$ , then

$$\sum_{j=1}^{\ell} z''_{i_j} \leq \sum_{j=1}^{\ell} y_{k+j}.$$

For the second, if  $z \in C$ , then

$$\sum_{j=1}^k z_j + \sum_{j=1}^{\ell} z_{i_j+k} \leq \sum_{j=1}^k y_j + \sum_{j=k+1}^{\ell} y_j = \sum_{j=1}^{k+\ell} y_j.$$

Since  $\sum_{j=1}^k z_i = \sum_{j=1}^k y_j$ , this implies that

$$\sum_{j=1}^{\ell} z''_{i_\ell} = \sum_{j=1}^{\ell} z_{i_\ell+k} \leq \sum_{j=k+1}^{k+\ell} y_j,$$

as desired. For the first, use the same trick to notice that

$$\sum_{j=1}^k y_j + \sum_{j=1}^{n-k} z''_j + \sum_{j=1}^k z_i + \sum_{j=1}^{n-k} z''_j = \sum_{j=1}^n z_j = \sum_{j=1}^n y_j.$$

Finally, suppose  $z$  is extreme in  $C$  (still supposing  $\sum_{i=1}^k z_i = \sum_{i=1}^k y_i$ ). Suppose  $z'_1, z'_2 \in C'$ ,  $t \in (0, 1)$  and  $(1-t)z'_1 + tz'_2 = z'$ . We need to show that  $z'_1 = z'_2 = z'$ . Set  $z_1 = (z'_1, z'')$ ,  $z_2 = (z'_2, z'')$ . We have shown that  $z'' \in C''$ , and so by the first part of this lemma,  $z_1, z_2 \in C$ , and by definition  $(1-t)z_1 + tz_2 = z$ . Since  $z$  is extreme,  $z_1 = z_2 = z$  and so consequently,  $z'_1 = z'_2 = z'$ . The same argument shows that  $z''$  is extreme.  $\square$

We continue with the inductive proof. Assume the claim for  $m < n$ . We prove it for  $n$ . Let  $e$  be an extreme point of  $C$ . Since  $C$  is invariant under permutation, we may assume that  $e = e^{\downarrow}$ . Let  $k$  is the smallest integer at least 1 for which  $\sum_{j=1}^k e_j = \sum_{j=1}^k y_j$  (by definition this holds for  $k = n$ , so  $k$  is well-defined).

Write  $e = (e', e'')$ , where  $e'$  comprises the first  $k$  components and  $e''$  comprises the last  $n-k$  components. By the lemma,  $e' \in C'$  and  $e'' \in C''$  ( $C'$ ,  $C''$  defined as in the lemma) are both extreme. Since  $k \geq 1$ , by induction  $e''$  has the form

$$(y_{\sigma(1)+k}, \dots, y_{\sigma(n-k)+k})$$

(or  $k = n$  in which case there is no  $e''$ ), for some  $\sigma \in S_{n-k}$ . Since  $y = y^{\downarrow}$  and  $e = e^{\downarrow}$ , we can assume without loss of generality that  $\sigma$  is the identity. If we can show  $k = 1$ , then since  $e'$  is extreme, by induction  $e' = (y_1)$ , and thus

$$e = y \subseteq \{y^\sigma : \sigma \in S_n\}$$

as desired.<sup>3</sup>

Suppose not, and  $k > 1$ . Then by definition of  $k$ , for  $1 \leq \ell < k$

$$(1) \quad \sum_{j=1}^{\ell} e_j < \sum_{j=1}^{\ell} y_j.$$

For  $\varepsilon > 0$  sufficiently small (which we choose later), define

$$e'^{\pm} = (e_1 \pm \varepsilon, e_2, \dots, e_{k-1}, e_k \mp \varepsilon).$$

We show that  $e'^{\pm} \in C'$ . Now suppose  $1 \leq \ell \leq k$  and  $i_1, \dots, i_{\ell}$  are the indices of the  $\ell$  largest elements of  $e'^{\pm}$ . We need to show that

$$\sum_{j=1}^{\ell} e'_{i_j}^{\pm} \leq \sum_{j=1}^{\ell} y_{i_j},$$

with equality if  $\ell = k$  (i.e. the sum runs from  $j = 1, \dots, k$ ). If  $\ell = k$ , then

$$\sum_{j=1}^k e'_{i_j}^{\pm} = \sum_{j=1}^k l e'_j{}^{\pm} = \sum_{j=1}^k e'_j{}^{\pm} \pm \varepsilon \mp \varepsilon = \sum_{j=1}^k y_j.$$

If  $\ell < k$ , then for  $\varepsilon$  sufficiently small, by (1)

$$\sum_{j=1}^{\ell} e'_{i_j}^{\pm} \leq \sum_{j=1}^{\ell} e'_{i_j} + \varepsilon \leq \sum_{i=1}^{\ell} e'_i + \varepsilon < \sum_{j=1}^{\ell} y_j$$

Thus  $e'^{\pm} \in C'$ . Set  $e^{\pm} = (e'^{\pm}, e'')$ . Recall that  $e'' \in C''$ , and so by the lemma,  $e^{\pm} \in C$ . By definition  $e = \frac{1}{2}e^- + \frac{1}{2}e^+$ , which contradicts extremity. Thus  $k = 1$  and the proof is complete.  $\square$

## 2. APPLICATION TO FUNCTIONAL ANALYSIS

Now we can give an example application to functional analysis.

**Theorem 2.1.** *Suppose  $A$  and  $B$  are compact positive operators on a Hilbert space,  $H$ . Suppose that  $A \leq B$ , or more generally that  $\text{Tr}(\Pi A) \leq \text{Tr}(\Pi B)$  for any orthogonal projection onto a finite-dimensional subspace. Then if  $\varphi$  is any non-negative convex non-decreasing function on  $\mathbf{R}$  such that  $\text{Tr}(\varphi(B)) < \infty$ , then  $\text{Tr}(\varphi(A)) \leq \text{Tr}(\varphi(B))$ .*

This and similar theorems have important applications in the theory of the Schatten class,  $\ell^p(H)$ , of operators  $A$  with  $\text{Tr}(|A|^p) < \infty$ . We prove this theorem in a few parts. First, notice that if  $A \leq B$  and  $\Pi$  is a projection onto a finite-dimensional subspace of  $H$  with orthonormal basis  $\{e_1, \dots, e_n\}$ , then

$$\text{Tr}(\Pi A) = \sum_{i=1}^n \langle A e_i, e_i \rangle \leq \sum_{i=1}^n \langle B e_i, e_i \rangle = \text{Tr}(\Pi B),$$

so that the assumption  $A \leq B$  really is a special case of the more general assumption. The main tool we use is the following inequality, occasionally attributed to Ky Fan. If  $\Sigma \subseteq H$  is finite-dimensional, then write  $\Pi_{\Sigma}$  for the orthogonal projection onto  $\Sigma$ .

<sup>3</sup>Recall that we are assuming  $e = e^{\downarrow}$ , so we did not expect  $e = y^{\sigma}$  for any non-identity  $\sigma$ .

**Proposition 2.2.** *Suppose  $C$  is a positive self-adjoint operator on a Hilbert space,  $H$ , with eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots$ . Then*

$$\max_{\dim(\Sigma)=k} \text{Tr}(\Pi_\Sigma C) = \sum_{i=1}^k \lambda_i.$$

With this proposition, the proof of the theorem is easy.

*Proof of theorem 2.1.* If  $\lambda_1 \geq \lambda_2 \geq \dots$  and  $\mu_1 \geq \mu_2 \geq \dots$  are the eigenvalues of  $A$  and  $B$ , respectively, then the hypotheses and the proposition give that for any  $k$

$$\sum_{i=1}^k \lambda_i \leq \sum_{i=1}^k \mu_i.$$

Karamata's inequality now yields the result that for any  $n$

$$\sum_{i=1}^n \varphi(\lambda_i) \leq \sum_{i=1}^n \varphi(\mu_i).$$

Suppose  $e_1, e_2, \dots$  is an orthonormal basis of  $H$  consisting of eigenvectors for  $A$ , ordered so that with  $Ae_i = \lambda_i e_i$ , then  $\lambda_i$  is non-increasing. Then taking limits, it follows that

$$\text{Tr}(\varphi(A)) = \sum_{i=1}^{\infty} \langle \varphi(A)e_i, e_i \rangle = \sum_{i=1}^{\infty} \varphi(\lambda_i) \leq \sum_{i=1}^{\infty} \varphi(\mu_i) = \text{Tr}(\varphi(B)).$$

□

Now we prove the proposition.

*Proof of proposition 2.2.* Suppose  $e_1, e_2, \dots$  is an orthonormal basis of  $H$  consisting of eigenvectors of  $C$ , ordered so that with  $Ce_i = \lambda_i e_i$ ,  $\lambda_i$  is non-increasing. Then, with  $\Sigma = \text{span}\{e_1, \dots, e_k\}$ , certainly

$$\text{Tr}(\Pi_\Sigma C) = \sum_{i=1}^k \lambda_i.$$

To prove the proposition, we just need to show that if  $\dim \Sigma = k$  is any other subspace, then  $\text{Tr}(\Pi_\Sigma C) \leq \sum_{i=1}^k \lambda_i$ . Write  $a_i = \|\Pi_\Sigma e_i\|^2 = \langle \Pi_\Sigma e_i, e_i \rangle$ . Then  $0 \leq a_i \leq 1$ . Expanding in the basis  $e_1, e_2, \dots$ ,

$$\begin{aligned} \text{Tr}(\Pi_\Sigma C) &= \sum_{i=1}^{\infty} \lambda_i a_i = \sum_{i=1}^k \lambda_i + \sum_{i=1}^k \lambda_i (a_i - 1) + \sum_{i=k+1}^{\infty} \lambda_i a_i \\ &\leq \sum_{i=1}^k \lambda_i + \sum_{i=1}^k \lambda_k (a_i - 1) + \sum_{i=k+1}^{\infty} \lambda_k a_i \\ &= \sum_{i=1}^k \lambda_i - k\lambda_k + \lambda_k \sum_{i=1}^{\infty} a_i. \end{aligned}$$

The manipulations with the infinite sums are justified since  $\Pi_\Sigma C$  has finite rank, and is thus trace class. Alternatively, one can just replace the  $\infty$  with some large  $N$  in the computation, and take limits. Now, if  $f_1, \dots, f_k, f_{k+1}, \dots$  is an orthonormal basis of  $H$  such that  $\{f_1, \dots, f_k\}$  is an orthonormal basis of  $\Sigma$ , then

$$\sum_{i=1}^{\infty} a_i = \sum_{i=1}^{\infty} \langle \Pi_\Sigma e_i, e_i \rangle = \sum_{i=1}^k \langle \Pi_\Sigma f_i, f_i \rangle = k.$$

Therefore,

$$\text{Tr}(\Pi_\Sigma C) \leq \sum_{i=1}^k \lambda_i - k\lambda_k + k\lambda_k = \sum_{i=1}^k \lambda_i,$$

which is the desired inequality.  $\square$

## APPENDIX A. APPENDIX

**Theorem A.1** (Finite-dimensional Krein-Milman theorem). *Let  $K \subseteq \mathbf{R}^n$  be compact and convex. Then  $K$  is the convex hull of its extreme points.*

*Proof.* We use induction on  $n$ . If  $n = 1$ ,  $K = [a, b]$  and the theorem is obvious. Assume the theorem now for  $n - 1$ ; we prove it for  $n$ . Let  $x \in K$ . We need to show that  $x$  is a convex combination of extreme points in  $K$ . Fix any  $v$  with  $\|v\| = 1$ , and consider the line  $s \mapsto x + sv$ . Since  $K$  is compact and convex there exist finite minimal and maximal times,  $s^- \leq 0 \leq s^+$ , for which  $x^\pm := x + s^\pm v \in K$  and  $x + sv \notin K$  for  $s > s^+$  or  $s < s^-$ . We will show that  $x^\pm$  are both convex combinations of extreme points of  $K$ , and thus since  $x$  lies on the line segment between  $x^-$  and  $x^+$ , it is a convex combination of  $x^-$  and  $x^+$ , and hence also a convex combination of extreme points of  $K$ .

We focus on the case for  $x^+$ . The case of  $x^-$  is handled in the same way. We show that there is some unit vector  $w \in \mathbf{R}^n$  such that  $\langle x^+, w \rangle$  maximizes the functional  $\varphi = \langle \bullet, w \rangle$  over  $K$ . By definition,  $x^+ + 1/n \notin K$  for all  $n \in \mathbf{N}$ . In particular, we can separate  $x^+ + 1/n$  from  $K$  with a hyperplane, i.e. there exists a unit vector  $w_n \in \mathbf{R}^n$  such that

$$(2) \quad \langle x^+ + 1/n, w_n \rangle \geq \langle z, w_n \rangle$$

for all  $z \in K$ . Since the  $w_n$  are bounded, we may extract a convergent subsequence  $w_{n_k} \rightarrow w$ . Taking limits in (2) yields

$$\langle x^+, w \rangle \geq \langle z, w \rangle$$

for all  $z \in K$ , i.e.  $x^+$  maximizes  $\varphi$  over  $K$ . Set  $a = \langle x^+, w \rangle$ , and set

$$K' = (K \cap \varphi^{-1}(a)) - x^+ = (K - x^+) \cap \varphi^{-1}(0).$$

Notice that  $K'$  is compact and convex, and contains 0. Since  $\varphi$  is a nonzero functional,  $\dim \ker \varphi = n - 1$ , and  $\ker \varphi$  is isomorphic to  $\mathbf{R}^{n-1}$  as a vector space.<sup>4</sup> Identifying  $K'$  with its image under this isomorphism, by induction every point in  $K'$  is a convex combination of its extreme points. In particular 0 is a convex combination of the extreme points of  $K'$ . We will show that for every extreme point  $e \in K'$ ,  $e + x^+$ , which is in  $K$ , is an extreme point of  $K$ . Thus if

$$0 = t_1 e_1 + \dots + t_k e_k$$

---

<sup>4</sup>Notice that notions of convexity and being an extreme point, etc., depend only on the vector-space structure, so are carried through under the isomorphism.



is a convex combination of extreme points in  $K'$ , then

$$x^+ = t_1(e_1 + x^+) + \cdots + t_k(e_k + x^+)$$

is a convex combination of extreme points in  $K$ , which is what we wanted to show.

Suppose  $e \in K'$  is extreme, and  $e + x^+ = (1-t)z_1 + tz_2$  for  $z_1, z_2 \in K$  and  $t \in (0, 1)$ . Recall that  $a = \varphi(x^+)$  is the maximum of  $\varphi$  over  $K$ . Then

$$a = \varphi(e + x^+) = (1-t)\varphi(z_1) + t\varphi(z_2) \leq (1-t)a + ta = a.$$

Thus equality must hold throughout, and  $\varphi(z_1) = \varphi(z_2) = a$ , and hence  $z_1 - x^+, z_2 - x^+ \in K'$ . Since  $e = (1-t)(z_1 - x^+) + t(z_2 - x^+)$  and  $e$  is extreme in  $K'$ , this means  $z_1 - x^+ = z_2 - x^+ = e$ . Thus  $e + x^+$  is extreme in  $K$ , as desired.  $\square$