

# **Robust Weighting and Matching Techniques for Causal Inference in Observational Studies with Continuous Treatment**

Universität Stuttgart



Universität Stuttgart

Ioan Scheffel

December 17, 2022

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Balancing Weights</b>	<b>4</b>
2.1	Introduction . . . . .	4
2.2	Estimating the Population Mean of Potential Outcomes . . . . .	4
2.3	Application of Convex Optimization . . . . .	4
2.4	Application of Matrix Concentration Inequalities . . . . .	5
<b>3</b>	<b>Convex Analysis</b>	<b>6</b>
3.1	Basic Notions . . . . .	6
3.2	Relative interiors . . . . .	6
3.3	Convex Separation . . . . .	7
3.4	Relative Interior . . . . .	8
3.5	Conjugate Calculus . . . . .	9
<b>4</b>	<b>Random Matrix Inequalities</b>	<b>12</b>
4.1	Matrix Analysis . . . . .	12
4.2	Matrix Khintchin Inequality . . . . .	12
4.3	Matrix Moment Inequality . . . . .	13
4.4	Intrinsic Dimension . . . . .	13
<b>5</b>	<b>Empirical Processes</b>	<b>14</b>
<b>6</b>	<b>Simple yet useful Calculations</b>	<b>15</b>

# 1 Introduction

## 2 Balancing Weights

### 2.1 Introduction

### 2.2 Estimating the Population Mean of Potential Outcomes

### 2.3 Application of Convex Optimization

**Assumption 2.1.** Assume that the map  $f : \mathbb{R} \rightarrow \overline{\mathbb{R}}$  has the following properties.

- (i)  $f$  is strictly convex.
- (ii)  $f$  is lower-semicontinuous and continuously differentiable on  $\text{int}(\text{dom}(f))$ .
- (iii) The derivative of  $f$  on  $\text{int}(\text{dom}(f))$  is a diffeomorphism.
- (iv) The Legendre transformation  $f^*$  of  $f$  is finite.
- (v) The function  $x \mapsto xt - f(x)$  takes its supremum on  $\text{int}(\text{dom}(f))$  for all  $t \in \mathbb{R}$ .

We consider the following optimization problem.

**Problem 2.1.**

$$\underset{w_1, \dots, w_n \in \mathbb{R}}{\text{minimize}} \quad \sum_{i=1}^n T_i f(w_i)$$

subject to the constraints

$$\begin{aligned} w_i T_i &\geq 0, & i &= 1, \dots, n, \\ \sum_{i=1}^n w_i T_i &= 1 \\ \left| \sum_{i=1}^n w_i T_i B_k(X_i) - \frac{1}{n} \sum_{i=1}^n B_k(X_i) \right| &\leq \delta_k, & k &= 1, \dots, K \end{aligned}$$

**Theorem 2.1.** *Under Assumption, the dual of the above Problem is the unconstrained optimization problem*

$$\underset{\lambda \in \mathbb{R}^K}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n nT_i f^*(\langle B(X_i), \lambda \rangle) - \langle B(X_i), \lambda \rangle + \langle \delta, |\lambda| \rangle,$$

where  $t \mapsto f^*(t) = t(f')^{-1}(t) - f((f')^{-1}(t))$  is the Legendre transformation of  $f$ ,  $B(X_i) = (B_1(X_i), \dots, B_K(X_i))$  denotes the  $K$  basis functions of the covariates of unit  $i \in \{1, \dots, n\}$  and  $|\lambda| = (|\lambda_1|, \dots, |\lambda_K|)$ , where  $|\cdot|$  is the absolute value of a real-valued scalar. Moreover, if  $\lambda^\dagger$  is an optimal solution then

$$w_i^* = (f')^{-1}(\langle B(X_i), \lambda^\dagger \rangle), \quad i \in \{1, \dots, n\} \quad (2.1)$$

are the unique optimal solutions to (P) .

**Proof.** We begin by proving the following Lemma.

**Lemma 2.1.** *The dual of the optimization problem is*

$$\underset{\lambda \in \mathbb{R}^{2K}}{\text{minimize}} \quad \frac{1}{n} \sum_{i=1}^n nT_i f^*(\langle Q_i, \lambda \rangle) - \langle Q_i, \lambda \rangle + \langle d, \lambda \rangle$$

subject to

$$\lambda_k \geq 0 \quad \text{for all } k \in \{1, \dots, K\}, \quad (2.2)$$

where

$$\mathbf{Q} := (\mathbf{B}(\mathbf{X}), -\mathbf{B}(\mathbf{X})) \quad (2.3)$$

□

## 2.4 Application of Matrix Concentration Inequalities

# 3 Convex Analysis

## 3.1 Basic Notions

Excursively, we present some well known definitions and facts from convex analysis. For details, see, e.g., [MMN22].

A subset  $C \subseteq \mathbb{R}^n$  is called **convex set**, if for all  $x, y \in C$  and all  $\lambda \in [0, 1]$ , we have  $\lambda x + (1 - \lambda)y \in C$ . The Cartesian product of convex sets is convex. The intersection of a collection of convex sets is also convex. Given (not necessary convex) sets  $\Omega, \Omega_1, \Omega_2 \subseteq \mathbb{R}^n$  and  $\lambda \in \mathbb{R}$ , define the **set addition** and **multiplication** by a real scalar as  $\Omega_1 + \Omega_2 := \{x_1 + x_2 : x_1 \in \Omega_1, x_2 \in \Omega_2\}$  and  $\lambda\Omega := \{\lambda x : x \in \Omega\}$ . For convex sets the addition and multiplication by a real scalar are convex.

A mapping  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is called **affine mapping** if there exist a linear mapping  $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$  and a vector  $b \in \mathbb{R}^m$  such that  $A(x) = L(x) + b$  for all  $x \in \mathbb{R}^n$ . The image and inverse image/preimage of convex sets under affine mappings are also convex.

## 3.2 Relative interiors

**Definition 3.1.** (Affine set and hull) *A set  $A \subseteq \mathbb{R}^n$  is called **affine**, if*

$$\alpha x + (1 - \alpha)y \in A \quad \text{for all } x, y \in A \text{ and } \alpha \in \mathbb{R}. \quad (3.1)$$

*The **affine hull**  $\text{aff}(\Omega)$  of a set  $\Omega \subseteq \mathbb{R}^n$  is the smallest affine set that includes  $\Omega$ .*

**Definition 3.2.** (Relative interior) *Let  $\Omega \subseteq \mathbb{R}^n$ . We define the **relative interior** of  $\Omega$  by*

$$\text{ri}(\Omega) := \{x \in \Omega : \text{there exists } \gamma > 0 \text{ such that } B_\gamma(x) \cap \text{aff}(\Omega) \subseteq \Omega\}. \quad (3.2)$$

**Proposition 3.1.** *Let  $C$  be a non-empty convex set in  $\mathbb{R}^n$ . Then we get the representation*

$$\text{ri}(C) = \{z \in C : \text{for all } x \in C \text{ there exists } t > 0 \text{ such that } z + t(z - x) \in C\}. \quad (3.3)$$

**Proof.** [Roc70, Theorem 6.4] □

**Proposition 3.2.** *Let  $C_1 \subseteq \mathbb{R}^{n_1}$  and  $C_2 \subseteq \mathbb{R}^{n_2}$  be two non-empty convex sets. Then it holds*

$$\text{ri}(C_1 \times C_2) = \text{ri}(C_1) \times \text{ri}(C_2). \quad (3.4)$$

**Proof.** Let  $(z_1, z_2) \in \text{ri}(C_1 \times C_2)$ . Then for all  $(x_1, x_2) \in C_1 \times C_2$  there exists  $t > 0$  such that

$$z_i + t(z_i - x_i) \in C_i \quad \text{for } i \in \{1, 2\}. \quad (3.5)$$

This proves  $\subseteq$ . Suppose  $z_1 \in \text{ri}(C_1)$  and  $z_2 \in \text{ri}(C_2)$ . Let  $(x_1, x_2) \in C_1 \times C_2$  with corresponding  $t_1, t_2 > 0$ . If  $t_1 = t_2$  everything is clear. W.l.o.g. assume  $t_1 < t_2$ . Define  $\theta := \frac{t_1}{t_2} \in (0, 1)$ . By the convexity of  $C_2$  it follows

$$z_2 + t_1(z_2 - x_2) = \theta(z_2 + t_2(z_2 - x_2)) + (1 - \theta)z_2 \in C_2. \quad (3.6)$$

Thus  $(z_1, z_2) \in \text{ri}(C_1 \times C_2)$ . This proves  $\supseteq$  and equality. □

### 3.3 Convex Separation

**Definition 3.3.** Let  $C_1$  and  $C_2$  be two non-empty convex sets in  $\mathbb{R}^n$ . A hyperplane  $H$  is said to **separate**  $C_1$  and  $C_2$  if  $C_1$  is contained in one of the closed half-spaces associated with  $H$  and  $C_2$  lies in the opposite closed half-space. It is said to separate  $C_1$  and  $C_2$  **properly** if  $C_1$  and  $C_2$  are not both actually contained in  $H$  itself.

**Theorem 3.1.** Let  $C_1$  and  $C_2$  be two non-empty convex sets in  $\mathbb{R}^n$ . There exists a hyperplane separating  $C_1$  and  $C_2$  properly if and only if there exists a vector  $b \in \mathbb{R}^n$  such that

$$\sup_{x \in C_2} \langle x, b \rangle \leq \inf_{x \in C_1} \langle x, b \rangle \quad \text{and} \quad \inf_{x \in C_2} \langle x, b \rangle < \sup_{x \in C_1} \langle x, b \rangle. \quad (3.7)$$

**Proof.** [Roc70, Theorem 11.1] □

**Theorem 3.2.** (Convex separation in finite dimension) Let  $C_1$  and  $C_2$  be two non-empty convex sets in  $\mathbb{R}^n$ . Then  $C_1$  and  $C_2$  can be properly separated if and only if  $\text{ri}(C_1) \cap \text{ri}(C_2) = \emptyset$ .

**Proof.** [Roc70, Theorem 11.3] □

**Definition 3.4.** (Support function intersection rule) (Support function) Given a nonempty subset  $\Omega \subseteq \mathbb{R}^n$  the **support function**  $\sigma_\Omega : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  of  $\Omega$  is defined by

$$\sigma_\Omega(x^*) := \sup_{x \in \Omega} \langle x^*, x \rangle \quad \text{for } x^* \in \mathbb{R}^n. \quad (3.8)$$

**Theorem 3.3.** Let  $C_1$  and  $C_2$  be two non-empty convex sets in  $\mathbb{R}^n$  with  $\text{ri}(C_1) \cap \text{ri}(C_2) \neq \emptyset$ . Then the support function of the intersection  $C_1 \cap C_2$  is represented as

$$(\sigma_{C_1 \cap C_2})(x^*) = (\sigma_{C_1} \square \sigma_{C_2})(x^*) \quad \text{for all } x^* \in \mathbb{R}^n. \quad (3.9)$$

Furthermore, for any  $x^* \in \text{dom}(\sigma_{C_1 \cap C_2})$  there exist dual elements  $x_1^*, x_2^* \in \mathbb{R}^n$  such that  $x^* = x_1^* + x_2^*$ . and

$$(\sigma_{C_1 \cap C_2})(x^*) = \sigma_{C_1}(x_1^*) + \sigma_{C_2}(x_2^*). \quad (3.10)$$

**Proof.** [MMN22, Theorem 4.23] We define

$$\Theta_1 := C_1 \times [0, \infty) \quad \text{and} \quad \Theta_2 := \{(x, \lambda) \in \mathbb{R}^n \times \mathbb{R} : x \in C_2 \text{ and } \lambda \leq \langle x^*, x \rangle - \alpha\}. \quad (3.11)$$

Clearly  $\Theta_1$  is convex by the convexity of  $C_1$ . Consider the affine function

$$\varphi : \mathbb{R}^{n+1} \rightarrow \mathbb{R}, \quad (x, \lambda) \mapsto \alpha - \langle x^*, x \rangle - \lambda. \quad (3.12)$$

It holds  $\Theta_2 = \varphi^{-1}((-\infty, 0]) \cap (C_2 \times \mathbb{R})$ . Thus, by the convexity of the sets  $\varphi^{-1}((-\infty, 0])$  and  $C_2$  it follows that  $\Theta_2$  is convex. We want to apply convex separation to  $\Theta_1$  and  $\Theta_2$ . To this end we show  $\text{ri}(\Theta_1) \cap \text{ri}(\Theta_2) = \emptyset$ . First note that

$$\text{ri}(\Theta_1) = \text{ri}(C_1) \times \text{ri}([0, \infty)) \subseteq \text{ri}(C_1) \times (0, \infty). \quad (3.13)$$

Indeed, if  $0 \in \text{ri}([0, \infty))$  then there exists  $t > 0$  such that  $-tx \geq 0$  for some  $x > 0$ . A contradiction. Furthermore

$$\text{ri}(\Theta_2) \subseteq \{(x, \lambda) \in \mathbb{R}^n : x \in \text{ri}(C_2) \text{ and } \lambda < \langle x^*, x \rangle - \alpha\}. \quad (3.14)$$

To see this, assume there is  $(x, \lambda) \in \text{ri}(\Theta_2)$  with  $\lambda = \langle x^*, x \rangle - \alpha$ . Then for some  $(y, \mu) \in \Theta_2$  with  $\mu < \langle x^*, y \rangle - \alpha$  there exists  $t > 0$  such that  $(x, \lambda) + t((x, \lambda) - (y, \mu)) \in \Theta_2$ . It follows

$$0 \leq (1+t)(\langle x^*, x \rangle - \alpha - \lambda) + t(\mu - \langle x^*, y \rangle + \alpha) < 0, \quad (3.15)$$

a contradiction. The first inequality is due to  $(x, \lambda) + t((x, \lambda) - (y, \mu)) \in \Theta_2$  and the second inequality due to  $\mu < \langle x^*, y \rangle - \alpha$  and  $\lambda = \langle x^*, x \rangle - \alpha$ . But then  $\text{ri}(\Theta_1) \cap \text{ri}(\Theta_2) = \emptyset$ . Indeed, suppose that there exists  $(x, \lambda) \in \text{ri}(\Theta_1) \cap \text{ri}(\Theta_2)$ . Then it holds  $\langle x^*, x \rangle - \alpha \leq 0$  and  $\lambda > 0$  since  $x \in \text{ri}(C_1) \cap \text{ri}(C_2) \subseteq C_1 \cap C_2$ . On the other hand

$$0 < \lambda < \langle x^*, x \rangle - \alpha \leq 0, \quad (3.16)$$

a contradiction.  $\square$

**Takeaways** Lorem ipsum dolor sit amet, consectetur adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetur id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

### 3.4 Relative Interior

**Definition 3.5.** Let  $\Omega \subseteq \mathbb{R}^n$ . We define the **relative interior** of  $\Omega$  by

$$\text{ri}(\Omega) := \{x \in \Omega : \text{there exists } \gamma > 0 \text{ such that } B_\gamma(x) \cap \text{aff}(\Omega) \in \Omega\}. \quad (3.17)$$

Next we collect some useful properties of relative interiors.

**Theorem 3.4.**

**Theorem 3.5.** Let  $C$  be a non-empty convex set in  $\mathbb{R}^n$ . Then we get the representation

(i)  $\text{ri}(C) = \{z \in C : \text{for all } x \in C \text{ there exists } t > 0 \text{ such that } z + t(z - x) \in C\}.$

(ii)  $\text{cl}(C)$  and  $\text{ri}(C)$  are convex sets.

(iii)  $\text{cl}(\text{ri}(C)) = \text{cl}(C)$  and  $\text{ri}(\text{cl}(C)) = \text{ri}(C).$

(iv) Suppose  $\bigcap_{i \in I} C_i \neq \emptyset$  for a finite index set  $I$ . Then  $\text{ri}(\bigcap_{i \in I} C_i) = \bigcap_{i \in I} \text{ri}(C_i).$

(v) Let  $L : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be a linear mapping. Then  $\text{ri}(L(C)) = L(\text{ri}(C)).$  If additionally it holds  $L^{-1}(\text{ri}(C)) \neq \emptyset$  we have  $\text{ri}(L^{-1}(C)) = L^{-1}(\text{ri}(C)).$

(vi)  $\text{ri}(C_1 \times C_2) = \text{ri}(C_1) \times \text{ri}(C_2).$



(vii)  $\text{ri}(C_1) \cap \text{ri}(C_2) = \emptyset$  if and only if  $0 \notin \text{ri}(C_1 - C_2)$ .

### 3.5 Conjugate Calculus

When studying different primal problems such as (??) we often turn to the dual instead. Therefore we need some reliable tools. Being able to compute specific convex conjugates is one tool required.

**Definition 3.6.** (Convex conjugate) *Given a function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$ , the **convex conjugate**  $f^* : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  of  $f$  is defined as*

$$f^*(x^*) := \sup_{x \in \mathbb{R}^n} (x^*)^T x - f(x) \quad (3.18)$$

Note that  $f$  in Definition 3.6 does not have to be convex. On the other hand, the convex conjugate is always convex:

**Proposition 3.3.** *Let  $f : \mathbb{R}^n \rightarrow (-\infty, \infty]$  be a proper function. Then its convex conjugate  $f^* : \mathbb{R}^n \rightarrow (-\infty, \infty]$  is convex.*

**Definition 3.7.** *Given a nonempty subset  $\Omega \subseteq \mathbb{R}^n$  the **support function**  $\sigma_\Omega : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  of  $\Omega$  is defined by*

$$\sigma_\Omega(x^*) := \sup_{x \in \Omega} \langle x^*, x \rangle \quad \text{for } x^* \in \mathbb{R}^n. \quad (3.19)$$

**Lemma 3.1.** *For any proper function  $f : \mathbb{R}^n \rightarrow \overline{\mathbb{R}}$  we have*

$$f^*(x^*) = \sigma_{\text{epi}(f)}(x^*, -1) \quad \text{for } x^* \in \mathbb{R}^n. \quad (3.20)$$

**Proof.** Let  $x^* \in \mathbb{R}^n$  and  $(x, \lambda) \in \text{epi}(f)$ . Then  $x \in \text{dom}(f)$  and  $f(x) \leq \lambda$ . Thus

$$\langle x^*, x \rangle - f(x) \geq \langle x^*, x \rangle - \lambda \quad \text{for all } (x, \lambda) \in \text{epi}(f). \quad (3.21)$$

On the other hand  $(x, f(x)) \in \text{epi}(f)$  for all  $x \in \text{dom}(f)$ . It follows

$$\langle x^*, x \rangle - f(x) \leq \sup_{(x, \lambda) \in \text{epi}(f)} \langle x^*, x \rangle - \lambda \quad \text{for all } x \in \text{dom}(f). \quad (3.22)$$

Taking the supremum in the last two displays yields

$$f^*(x^*) = \sup_{x \in \text{dom}(f)} \langle x^*, x \rangle - f(x) = \sup_{(x, \lambda) \in \text{epi}(f)} \langle x^*, x \rangle - \lambda \quad (3.23)$$

$$= \sup_{(x, \lambda) \in \text{epi}(f)} \langle (x^*, -1), (x, \lambda) \rangle = \sigma_{\text{epi}(f)}(x^*, -1). \quad (3.24)$$

□

**Proposition 3.4.**

**Theorem 3.6.** (Conjugate Chain Rule) *Let  $A : \mathbb{R}^m \rightarrow \mathbb{R}^n$  be a linear map (matrix) and  $g : \mathbb{R}^n \rightarrow (-\infty, \infty]$  a proper convex function. If  $\text{Im}(A) \cap \text{ri}(\text{dom}(g)) \neq \emptyset$  it follows*

$$(g \circ A)^*(x^*) = \inf_{y^* \in (A^*)^{-1}(x^*)} g^*(y^*). \quad (3.25)$$

Furthermore, for any  $x^* \in \text{dom}(g \circ A)^*$  there exists  $y^* \in (A^*)^{-1}(x^*)$  such that  $(g \circ A)^*(x^*) = g^*(y^*)$ .

**Definition 3.8.** (Infimal convolution) *Given functions  $f_i : \mathbb{R}^n \rightarrow (-\infty, \infty]$  for  $i = 1, \dots, n$  the **infimal convolution** of these functions is defined as*

$$(f_1 \square \dots \square f_m)(x) := \inf_{\substack{x_i \in \mathbb{R}^n \\ \sum_{i=1}^m x_i = x}} \sum_{i=1}^m f_i(x_i) \quad (3.26)$$

**Theorem 3.7.** *Let  $f, g : \mathbb{R}^n \rightarrow (-\infty, \infty]$  be proper convex functions and  $\text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(g)) \neq \emptyset$ . Then we have the conjugate sum rule*

$$(f + g)^*(x^*) = (f^* \square g^*)(x^*) \quad (3.27)$$

*for all  $x^* \in \mathbb{R}^n$ . Moreover, the infimum in  $(f^* \square g^*)(x^*)$  is attained, i.e., for any  $x^* \in \text{dom}(f + g)^*$  there exists vectors  $x_1^*, x_2^*$  for which*

$$(f + g)^*(x^*) = f^*(x_1^*) + g^*(x_2^*), \quad x^* = x_1^* + x_2^*. \quad (3.28)$$

**Proof.** Let  $x^* \in \mathbb{R}^n$  and fix  $x_1^*, x_2^* \in \mathbb{R}^n$  such that  $x^* = x_1^* + x_2^*$ . We get

$$\begin{aligned} f^*(x_1^*) + g^*(x_2^*) &= \sup_{x \in \mathbb{R}^n} \langle x_1^*, x \rangle - f(x) + \sup_{x \in \mathbb{R}^n} \langle x_2^*, x \rangle - g(x) \\ &\geq \sup_{x \in \mathbb{R}^n} \langle x_1^*, x \rangle - f(x) + \langle x_2^*, x \rangle - g(x) = \sup_{x \in \mathbb{R}^n} \langle x_1^* + x_2^*, x \rangle - (f(x) + g(x)) \\ &= \sup_{x \in \mathbb{R}^n} \langle x^*, x \rangle - (f + g)(x) = (f + g)^*(x^*) \end{aligned}$$

Taking the infimum over  $x_1^*, x_2^* \in \mathbb{R}^n$  in the above display gives  $(f^* \square g^*)(x^*) \geq (f + g)^*(x^*)$ . Let us prove now  $\leq$  under the condition  $\text{ri}(\text{dom}(f)) \cap \text{ri}(\text{dom}(g)) \neq \emptyset$ . The only case we need to consider is  $(f + g)^*(x^*) < \infty$ . Define two convex sets by

$$\Omega_1 := \{(x, \alpha, \beta) \in \mathbb{R}^{n+2} : \alpha \geq f(x)\} = \text{epi}(f) \times \mathbb{R}, \quad (3.29)$$

$$\Omega_2 := \{(x, \alpha, \beta) \in \mathbb{R}^{n+2} : \beta \geq g(x)\}. \quad (3.30)$$

Similar to Lemma we get the representation

$$(f + g)^*(x^*) = \sigma_{\Omega_1 \cap \Omega_2}(x^*, -1, -1). \quad (3.31)$$

Indeed, the only thing we need to verify is  $\text{dom}(f) \cap \text{dom}(g) = \text{dom}(f + g)$ . The inclusion  $\subseteq$  is clear. Assume towards a contradiction that  $(f + g)(x) < \infty$  and  $f(x) = \infty$ . Since  $g(x) > -\infty$  it holds

$$\infty = \infty + g(x) = f(x) + g(x) = (f + g)(x) < \infty. \quad (3.32)$$

This is a contradiction. The same holds for  $f$  and  $g$  reversed. It follows the inclusion  $\supseteq$  and equality. By the support function intersection rule there exist triples

$$(x_1^*, -\alpha_1, -\beta_1), (x_2^*, -\alpha_2, -\beta_2) \in \mathbb{R}^{n+2} \quad \text{such that} \quad (x^*, -1, -1) = (x_1^* + x_2^*, -(\alpha_1 + \alpha_2), -(\beta_1 + \beta_2)) \quad (3.33)$$

and

$$(f + g)^*(x^*) = \sigma_{\Omega_1 \cap \Omega_2}(x^*, -1, -1) = \sigma_{\Omega_1}(x_1^*, -\alpha_1, -\beta_1) + \sigma_{\Omega_2}(x_2^*, -\alpha_2, -\beta_2). \quad (3.34)$$

Next we show  $\beta_1 = \alpha_2 = 0$ . Suppose towards a contradiction that  $\beta_1 \neq 0$ . We fix  $(\bar{x}, \bar{\alpha}) \in \text{epi}(f)$ . Then

$$\sigma_{\Omega_1}(x_1^*, -\alpha_1, -\beta_1) = \sup_{(x, \alpha, \beta) \in \text{epi}(f) \times \mathbb{R}} \langle x^*, x \rangle - \alpha \alpha_1 - \beta \beta_1 \geq \sup_{\beta \in \mathbb{R}} \langle x^*, \bar{x} \rangle - \bar{\alpha} \alpha_1 - \beta \beta_1 = \infty. \quad (3.35)$$

This contradicts  $(f+g)^*(x^*) < \infty$ . In a similar fashion we can derive a contradiction for  $\alpha_2 \neq 0$ . Employing Lemma and taking into account the structures of the sets  $\Omega_1$  and  $\Omega_2$  this implies

$$(f+g)^*(x^*) = \sigma_{\Omega_1 \cap \Omega_2}(x^*, -1, -1) = \sigma_{\Omega_1}(x_1^*, -1, 0) + \sigma_{\Omega_2}(x_2^*, 0, -1) \quad (3.36)$$

$$= \sigma_{\text{epi}(f)}(x_1^*, -1) + \sigma_{\text{epi}(g)}(x_2^*, -1) = f^*(x_1^*) + g^*(x_2^*) \geq (f^* \square g^*)(x^*). \quad (3.37)$$

This finishes the proof. □

# 4 Random Matrix Inequalities

## 4.1 Matrix Analysis

The **trace** of a square matrix, denoted by  $\text{tr}$ , is the sum of its diagonal entries, i.e.  $\text{tr}(\mathbf{B}) = \sum_{j=1}^d b_{jj}$  for  $\mathbf{B} \in \mathbb{M}_d$ . The trace is unitarily invariant, i.e.  $\text{tr}(\mathbf{B}) = \text{tr}(\mathbf{Q}\mathbf{B}\mathbf{Q}^*)$  for all  $\mathbf{B} \in \mathbb{M}_d$  for all unitary  $\mathbf{Q} \in \mathbb{M}_d$ . In particular, the existence of an eigenvalue value decomposition shows that the trace of a Hermitian matrix equals the sum of its eigenvalues. Let  $f : I \rightarrow \mathbb{R}$  where  $I \subseteq \mathbb{R}$  is an interval. Consider a matrix  $\mathbf{A} \in \mathbb{H}_d$  whose eigenvalues are contained in  $I$ . We define the matrix  $f(\mathbf{A}) \in \mathbb{H}_d$  using an eigenvalue decomposition of  $\mathbf{A}$  :

$$f(\mathbf{A}) = \mathbf{Q} \begin{bmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_d) \end{bmatrix} \mathbf{Q}^* \quad \text{where} \quad \mathbf{A} = \mathbf{Q} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix} \mathbf{Q}^*. \quad (4.1)$$

The definition of  $f(\mathbf{A})$  does not depend on which eigenvalue decomposition we choose. Any matrix function that arises in this fashion is called a **standard matrix function**.

**Proposition 4.1.** *Let  $f, g : I \rightarrow \mathbb{R}$  be real-valued functions on an interval  $I \subseteq \mathbb{R}$ , and let  $\mathbf{A} \in \mathbb{H}_d$  be a Hermitian matrix whose eigenvalues are contained in  $I$ .*

- (i) *If  $\lambda$  is an eigenvalue of  $\mathbf{A}$ , then  $f(\lambda)$  is an eigenvalue of  $f(\mathbf{A})$ .*
- (ii)  *$f(a) \leq g(a)$  for all  $a \in I$  implies  $f(\mathbf{A}) \preceq g(\mathbf{A})$ .*

## 4.2 Matrix Khintchin Inequality

**Theorem 4.1.** [MJC<sup>+</sup>14, Corollary 7.3] *Suppose that  $p = 1$  or  $p \geq 3/2$ . Consider a finite sequence  $(\mathbf{Y}_k)_{k \geq 1}$  of independent, random, Hermitian matrices and a deterministic sequence  $(\mathbf{A}_k)_{k \geq 1}$  for which*

$$\mathbf{E}[\mathbf{Y}_k] = 0 \quad \text{and} \quad \mathbf{Y}_k^2 \preceq \mathbf{A}_k^2 \quad \text{almost surely for all } k \geq 1. \quad (4.2)$$

*Then*

$$\mathbf{E} \left[ \left\| \sum_{k \geq 1} \mathbf{Y}_k \right\|_{2p}^{2p} \right]^{1/(2p)} \leq \sqrt{p - \frac{1}{2}} \left\| \left( \sum_{k \geq 1} (\mathbf{A}_k^2 + \mathbf{E}[\mathbf{Y}_k^2]) \right)^{1/2} \right\|_{2p}. \quad (4.3)$$

In particular, when  $(\xi_k)_{k \geq 1}$  is an independent sequence of Rademacher random variables,

$$\mathbf{E} \left[ \left\| \sum_{k \geq 1} \xi_k \mathbf{A}_k \right\|_{2p}^{2p} \right]^{1/(2p)} \leq \sqrt{2p-1} \left\| \left( \sum_{k \geq 1} \mathbf{A}_k^2 \right)^{1/2} \right\|_{2p}. \quad (4.4)$$

### 4.3 Matrix Moment Inequality

**Theorem 4.2.** Assume  $n \geq 3$

(i) Suppose that  $p \geq 1$ , and fix  $r \geq p \vee 2 \log(n)$ . Consider a finite sequence  $(\mathbf{S}_k)_{k \geq 1}$  of independent, random, positive-semidefinite matrices with dimension  $n \times n$ . Then

$$\mathbf{E} \left[ \left\| \sum_{k \geq 1} \mathbf{S}_k \right\|^p \right]^{1/p} \leq \left[ \left\| \sum_{k \geq 1} \mathbf{E}[\mathbf{S}_k] \right\|^{1/2} + 2\sqrt{er} \mathbf{E}[\max_{k \geq 1} \|\mathbf{S}_k\|^p]^{1/(2p)} \right]^2. \quad (4.5)$$

(ii) Suppose that  $p \geq 2$ , and fix  $r \geq p \vee 2 \log(n)$ . Consider a finite sequence  $(\mathbf{Y}_k)_{k \geq 1}$  of independent, symmetric, random, self-adjoint matrices with dimension  $n \times n$ . Then

$$\mathbf{E} \left[ \left\| \sum_{k \geq 1} \mathbf{Y}_k \right\|^p \right]^{1/p} \leq \sqrt{er} \left\| \left( \sum_{k \geq 1} \mathbf{E}[\mathbf{Y}_k^2] \right)^{1/2} \right\| + 2er \mathbf{E}[\max_{k \geq 1} \|\mathbf{S}_k\|^p]^{1/p}. \quad (4.6)$$

### 4.4 Intrinsic Dimension

**Definition 4.1.** For a positive-semidefinite matrix  $\mathbf{S}$ , the *intrinsic dimension* is the quantity

$$\text{intdim}(\mathbf{A}) := \frac{\text{tr} \mathbf{A}}{\|\mathbf{A}\|}.$$

**Lemma 4.1.** (Intrinsic dimension) Let  $\varphi : [0, \infty) \rightarrow \mathbb{R}$  be a convex function with  $\varphi(0) = 0$ . For any positive-semidefinite matrix  $\mathbf{S}$  it holds that

$$\text{tr}(\varphi(\mathbf{S})) \leq \text{intdim}(\mathbf{S}) \cdot \varphi(\|\mathbf{S}\|).$$

**Proof.** [Tro15, Lemma 7.5.1] Since  $\varphi$  is convex on any interval  $[0, L]$  with  $L > 0$  and  $\varphi(0) = 0$ , it holds

$$\varphi(a) \leq \left(1 - \frac{a}{L}\right) \varphi(0) + \frac{a}{L} \varphi(L) = \frac{a}{L} \varphi(L) \quad \text{for all } a \in [0, L]. \quad (4.7)$$

Since  $\mathbf{S}$  is positive-semidefinite, the eigenvalues of  $\mathbf{S}$  fall in the interval  $[0, L]$ , where  $L = \|\mathbf{S}\|$ .

$$\text{tr}(\varphi(\mathbf{S})) = \sum_{i=1}^d \varphi(\lambda_i) \leq \frac{\sum_{i=1}^d \lambda_i}{\|\mathbf{S}\|} \varphi(\|\mathbf{S}\|) = \frac{\text{tr}(\mathbf{S})}{\|\mathbf{S}\|} \varphi(\|\mathbf{S}\|) = \text{intdim}(\mathbf{S}) \cdot \varphi(\|\mathbf{S}\|). \quad (4.8)$$

□

# 5 Empirical Processes

Let  $(\Omega, \mathcal{A}, \mathbf{P})$  be a probability space and  $(\mathcal{X}, \Sigma)$  a measurable space. Let  $X_j : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (\mathcal{X}, \Sigma)$ ,  $j = 1, \dots, n$  be independent and identically-distributed (i.i.d.) random variables with probability distribution  $\mathbf{P}_X$  and  $\mathcal{F}$  a family of measurable functions  $f : (\mathcal{X}, \Sigma) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ . Consider the map

$$f \mapsto G_n f := \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n f(X_i) - \mathbf{P}_X f \right), \quad (5.1)$$

where  $\mathbf{P}_X f := \int_{\mathcal{X}} f d\mathbf{P}_X$ . We call  $(G_n f)_{f \in \mathcal{F}}$  the empirical process indexed by  $\mathcal{F}$ . Furthermore

$$\|G_n f\|_{\mathcal{F}} := \sup_{f \in \mathcal{F}} |G_n f|. \quad (5.2)$$

**Lemma 5.1.** (Bernstein Inequality for Empirical Processes) *For any bounded, measurable function  $f$  it holds for all  $t > 0$*

$$\mathbf{P}(|G_n f| > t) \leq 2 \exp \left( -\frac{1}{4} \frac{t^2}{\mathbf{P}_X(f^2) + t \|f\|_{\infty} / \sqrt{n}} \right) \quad (5.3)$$

**Proof.** By the Markov inequality it holds for all  $\lambda > 0$

$$\mathbf{P}(G_n f > t) \leq e^{-\lambda t} \mathbf{E} \exp(\lambda G_n f) \quad (5.4)$$

□

**Lemma 5.2.** *For any finite class  $\mathcal{F}$  of bounded, measurable, square-integrable functions, with  $|\mathcal{F}|$  elements, it holds*

$$\mathbf{E} \|G_n f\|_{\mathcal{F}} \lesssim \max_{f \in \mathcal{F}} \frac{\|f\|_{\infty}}{\sqrt{n}} \log(1 + |\mathcal{F}|) + \max_{f \in \mathcal{F}} \|f\|_{\mathbf{P}, 2} \sqrt{\log(1 + |\mathcal{F}|)}. \quad (5.5)$$

## 6 Simple yet useful Calculations

**Theorem 6.1.** (Multivariate Taylor Theorem) *Let  $f \in C^2(\mathbb{R}^n, \mathbb{R})$ . Then for all  $x, \Delta \in \mathbb{R}^n$  there exists  $\xi \in [0, 1]$  such that it holds*

$$\begin{aligned} f(x + \Delta) = & f(x) + \sum_{i=1}^n \frac{\partial f(x)}{\partial x_i} \Delta_i + \sum_{\substack{i,j=1 \\ i \neq j}}^n \frac{\partial^2 f(x + \xi \Delta)}{\partial x_i \partial x_j} \Delta_i \Delta_j \\ & + \frac{1}{2} \sum_{i=1}^n \frac{\partial^2 f(x + \xi \Delta)}{\partial x_i^2} \Delta_i^2 \end{aligned} \quad (6.1)$$

**Corollary 6.1.1.** *Let  $f \in C^2(\mathbb{R})$ . Then for all  $a, x, \Delta \in \mathbb{R}^n$  there exist  $\xi \in [0, 1]$  such that it holds*

$$f(a^T(x + \Delta)) - f(a^T x) = f'(a^T x) \Delta^T a + \frac{1}{2} f''(a^T(x + \xi \Delta)) \Delta^T A \Delta, \quad (6.2)$$

where  $A := aa^T \in \mathbb{R}^{n \times n}$ .

**Proof.** By the chain rule we have for all  $a, x, \Delta \in \mathbb{R}^n$  and  $\xi \in [0, 1]$

$$\frac{\partial^2 f(a^T(x + \xi \Delta))}{\partial x_i \partial x_j} = f''(a^T(x + \xi \Delta)) a_i a_j. \quad (6.3)$$

Since  $A := aa^T$  is symmetric we have

$$\Delta^T A \Delta = 2 \sum_{\substack{i,j=1 \\ i \neq j}}^n a_i a_j \Delta_i \Delta_j + \sum_{i=1}^n a_i^2 \Delta_i^2. \quad (6.4)$$

Plugging (6.3) and (6.4) into (6.1) yields (6.2). □

**Proposition 6.1.** *For all  $x, y \in \mathbb{R}$  it holds*

$$|x + y| - |x| \geq -|y| \quad (6.5)$$

**Proof.** Checking all 6 combinations of  $x + y, x, y$  being nonnegative or negative yields the result. □

# Notation Index

$\#A$  cardinality of the set  $A$

$\mathbf{E}[X|Y]$  conditional expectation of the random variable  $X$  with respect to  $\sigma(Y)$

$\mathbf{E}[X]$  expectation of the random variable  $X$

$\mathbf{Var}[X]$  variance of the random variable  $X$

$\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$  extension of the real numbers

$\xrightarrow{\mathcal{D}}$  convergence of distributions

$\mathbf{P}$  generic probability measure

$\mathbf{P}_X = \mathbf{P} \circ X^{-1}$  distribution of the random variable  $X$

$\mathbb{R}$  set of real numbers

$x \vee y, x \wedge y, x^+, x^-$  maximum, minimum, positive part, negative part of real numbers

$X \sim \mu$  the random variable has distribution  $\mu$



# Bibliography

- [MJC<sup>+</sup>14] Lester Mackey, Michael I. Jordan, Richard Y. Chen, Brendan Farrell, and Joel A. Tropp. Matrix concentration inequalities via the method of exchangeable pairs. *The Annals of Probability*, 42(3), May 2014.
- [MMN22] Boris S. Mordukhovich and Nguyen Mau Nam. *Convex Analysis and Beyond: Volume I: Basic Theory*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, Cham, 2022.
- [Roc70] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970.
- [Tro15] Joel A. Tropp. An Introduction to Matrix Concentration Inequalities, January 2015.