

Projection integrals and testing for majorizing measures and comparison theorems

Nikita Zhivotovskiy¹

15 October 2026

In recent work [1], Pathak and Zhivotovskiy express Gaussian width through an exact projection integral and study minoration by statistical testing. Zadik [2] later independently obtained a version of this integral representation for finite sets and used it to give a Bayesian proof of the majorizing measures theorem.

In this note we apply the ideas of [1] in three directions. First, we give an alternative proof of the lower bound of Pathak and Zhivotovskiy [3]. This bound extends the Gaussian majorizing measures lower bound to more general laws. We then give elementary proofs of Gaussian-type comparison theorems and construct admissible partitions for convex hulls and Minkowski sums. More generally, in finite dimensions these tools construct admissible sequences of optimal order for compact sets, using only metric projections, averaging, and successive coverings by balls.

Throughout, norms are Euclidean, $G \sim N(0, I_d)$, and $X \sim \mu$ is centered with $\mathbb{E}\|X\| < \infty$. We write $w_\mu(T) = \mathbb{E} \sup_{t \in T} \langle X, t \rangle$ and $w(T) = w_G(T)$. The translation condition used below is

$$\mathcal{KL}(\mu_h \| \mu_k) \leq \kappa \|h - k\|^2 \quad (h, k \in \mathbb{R}^d), \quad \mu_h = \mathcal{L}(X + h), \quad \kappa > 0. \quad (\text{KL})$$

1 From the full projection integral to admissible partitions

Following Talagrand [4, Definitions 2.7.1, 2.7.3], for a bounded nonempty subset T of a Hilbert space, we define

$$\gamma_2(T) = \inf_{(\mathcal{A}_n)} \sup_{t \in T} \sum_{n \geq 0} 2^{n/2} \text{diam}(A_n(t)).$$

The infimum is over sequences of partitions with $\mathcal{A}_0 = \{T\}$ and $|\mathcal{A}_n| \leq 2^{2^n}$, such that each partition refines the preceding one: every cell of $\mathcal{A}_{n+1}$ is contained in a cell of $\mathcal{A}_n$. Here $A_n(t)$ is the unique cell containing t . Such sequences are called admissible, and their cost is the displayed supremum. The following extension of the Gaussian majorizing measures lower bound is due to Pathak and Zhivotovskiy [3, Theorem 1]. We give a short proof.

Theorem 1. *Under (KL), every bounded nonempty $T \subset \mathbb{R}^d$ satisfies*

$$\gamma_2(T) \leq C \sqrt{\kappa} w_\mu(T),$$

where C is universal.

In particular, the Gaussian translation formula, $\mathcal{KL}(N(h, I_d) \| N(k, I_d)) = \|h - k\|^2/2$, gives the lower bound in the Gaussian majorizing measures theorem [4, Theorem 2.10.1]. The projection decomposition of Pathak and Zhivotovskiy [1] applies to every centered integrable noise vector. We combine it directly with testing and admissible partitions.

1. *The projection decomposition.* Let $K = \overline{\text{conv}} T$. For every $z \in K$, the projection identity is

$$w_\mu(T) = \frac{1}{2} \int_0^\infty \mathbb{E} \|\Pi_K(z + \sigma X) - z\|^2 \frac{d\sigma}{\sigma^2}, \quad (1)$$

¹Department of Statistics, UC Berkeley. zhivotovskiy@berkeley.edu. This note uses ideas of [1, 3]. AI tools (GPT models up to Astra 6 and Claude Opus 5.5) assisted with computations and preparation of the text. The note stays within the “convex hull” of the results and tools in the cited works.

where $\Pi_C(x) = \arg \min_{u \in C} \|x - u\|^2$ is the Euclidean projection onto a nonempty closed convex set C . This is the decomposition of [1, Theorem 2.2 and (15b)] applied to $K - z$; see also [3, Lemma 2, (6)]. Its proof uses comparison of maximizers and integration in one variable. The same comparison shows that $\|\Pi_K(z + \sigma x) - z\|^2$ is nondecreasing in σ .²

We use the following consequence of this identity. Write $\mathcal{R}_z(s) = \mathbb{E}\|\Pi_K(z + sX) - z\|^2$. For $q > 1$ and positive scales with $s_{j+1} \leq s_j/q$, the intervals $[s_j, qs_j]$ have disjoint interiors. Monotonicity therefore gives

$$\sum_j \frac{\mathcal{R}_z(s_j)}{s_j} \leq \frac{q}{q-1} \sum_j \int_{s_j}^{qs_j} \mathcal{R}_z(v) \frac{dv}{v^2} \leq \frac{2q}{q-1} w_\mu(T). \quad (2)$$

2. Statistical testing: a change of measure inequality. Summing an elementary change of measure inequality over disjoint events gives the following bound of Fano type; cf. [5, §§15.1.2, 15.3].

Lemma 1. *Let $P_1, \dots, P_m, Q$ be probability laws on the same measurable space, where $m \geq 2$, and let $E_1, \dots, E_m$ be pairwise disjoint measurable events. If $m^{-1} \sum_i P_i(E_i) \geq 7/8$, then*

$$\log m \leq \frac{16}{m} \sum_{i=1}^m \mathcal{KL}(P_i \| Q). \quad (3)$$

Proof. The scalar inequality $av \leq a \log a - a + \exp(v)$, integrated with $a = dP/dQ$, gives $\mathbb{E}_P V \leq \mathcal{KL}(P \| Q) + \mathbb{E}_Q \exp(V) - 1$. Take $V = (\log \lambda) \mathbf{1}_E$ and use $(\lambda - 1)/\log \lambda \leq \lambda$ to obtain, for $\lambda > 1$,

$$P(E) \leq \lambda Q(E) + \frac{\mathcal{KL}(P \| Q)}{\log \lambda}. \quad (4)$$

We sum over the disjoint events and take $\lambda = \sqrt{m}$:

$$\frac{7}{8} \leq \frac{1}{m} \sum_i P_i(E_i) \leq \frac{1}{\sqrt{m}} + \frac{2}{m \log m} \sum_i \mathcal{KL}(P_i \| Q).$$

Since $m^{-1/2} \leq 3/4$, this proves (3). ■

Replacing X by $\sqrt{\kappa}X$ reduces to $\kappa = 1$ and multiplies the width by $\sqrt{\kappa}$. For a set $B \subset T$ with $\text{diam}(B) > 0$ and an integer $n \geq 0$, put

$$\mathcal{G}_B = \left\{ z \in B : \mathcal{R}_z(5 \cdot 2^{-n/2} \text{diam}(B)) < \frac{\text{diam}(B)^2}{512} \right\}. \quad (5)$$

(We use variants of this set in Section 3; see (22).) We claim that $\mathcal{G}_B$ is covered by at most $2^{2^n} - 1$ closed balls of radius $\text{diam}(B)/4$. Otherwise, successively choosing a point outside the preceding balls gives 2^{2^n} points $t_i \in \mathcal{G}_B$ at pairwise distance greater than $\text{diam}(B)/4$. Use

$$P_i = \mathcal{L}(t_i + 5 \cdot 2^{-n/2} \text{diam}(B)X), \quad Q = P_1.$$

The events $E_i = \{y : \|\Pi_K(y) - t_i\| \leq \text{diam}(B)/8\}$ are disjoint. The Markov inequality gives

$$P_i(E_i^c) \leq \frac{64}{\text{diam}(B)^2} \mathcal{R}_{t_i}(5 \cdot 2^{-n/2} \text{diam}(B)) < \frac{64}{512} = \frac{1}{8}.$$

²Importantly, convexity is not strictly needed for either fact: for any nonempty compact set, we may use a fixed measurable nearest point selector.

Lemma 1 and (KL) now give the contradiction

$$2^n \log 2 \leq 16 \max_i \mathcal{KL}(P_i \| P_1) \leq \frac{16 \operatorname{diam}(B)^2}{25 \cdot 2^{-n} \operatorname{diam}(B)^2} = \frac{16}{25} 2^n < 2^n \log 2.$$

Thus the claimed cover exists.

3. *The admissible partitions.* Without loss of generality assume $\operatorname{diam}(T) > 0$. We start with $\mathcal{A}_0 = \{T\}$. Having defined $\mathcal{A}_n$, take a cell $B \in \mathcal{A}_n$ of positive diameter and form $\mathcal{G}_B$ as above at scale $5 \cdot 2^{-n/2} \operatorname{diam}(B)$. Choose closed balls $C_1, \dots, C_\ell$ of radius $\operatorname{diam}(B)/4$ covering $\mathcal{G}_B$, where $\ell \leq 2^{2^n} - 1$. In $\mathcal{A}_{n+1}$, replace B by the nonempty sets among

$$(B \cap C_i) \setminus \bigcup_{j < i} C_j \quad (1 \leq i \leq \ell), \quad B \setminus \bigcup_{i=1}^{\ell} C_i.$$

These sets are pairwise disjoint, have union B , and are called its children. Each of the first ℓ sets is contained in a ball C_i , so has diameter at most $\operatorname{diam}(B)/2$; the uncovered remainder lies in $B \setminus \mathcal{G}_B$. Children of distinct cells of $\mathcal{A}_n$ are disjoint as well. Thus applying this rule to every cell of positive diameter and retaining each singleton defines a partition $\mathcal{A}_{n+1}$ refining $\mathcal{A}_n$. Each cell has at most 2^{2^n} children, so

$$|\mathcal{A}_N| \leq \prod_{n < N} 2^{2^n} = 2^{2^N - 1} \leq 2^{2^N}. \quad (6)$$

Thus the sequence is admissible. Now fix $z \in T$ and put

$$a_n = 2^{n/2} \operatorname{diam}(A_n(z)), \quad s_n = 5 \cdot 2^{-n/2} \operatorname{diam}(A_n(z)).$$

If $A_{n+1}(z)$ is one of the children assigned to a covering ball C_i , it is contained in that ball. Its diameter is therefore at most half that of $A_n(z)$, giving $\sqrt{2}a_{n+1} \leq a_n$. If it is the uncovered remainder, then $z \notin \mathcal{G}_B$ for the cell $B = A_n(z)$. Thus $\mathcal{R}_z(s_n) \geq \operatorname{diam}(A_n(z))^2/512$, and

$$a_n = \frac{5 \operatorname{diam}(A_n(z))^2}{s_n} \leq 2560 \frac{\mathcal{R}_z(s_n)}{s_n}. \quad (7)$$

The inclusion $A_{n+1}(z) \subseteq A_n(z)$ also gives $\operatorname{diam}(A_{n+1}(z)) \leq \operatorname{diam}(A_n(z))$, hence $\sqrt{2}a_{n+1} \leq 2a_n$. Using (7) to bound one of these two copies of a_n , we obtain, in either case,

$$\sqrt{2}a_{n+1} \leq a_n + 2560 \frac{\mathcal{R}_z(s_n)}{s_n}. \quad (8)$$

Since $s_{n+1} \leq s_n/\sqrt{2}$, (2) with $q = \sqrt{2}$ gives

$$\sum_{n: s_n > 0} \frac{\mathcal{R}_z(s_n)}{s_n} \leq (4 + 2\sqrt{2})w_\mu(T).$$

For every N with $a_N > 0$, summing (8) over $0 \leq n \leq N$ and collecting the a_n terms gives

$$(\sqrt{2} - 1) \sum_{n=0}^N a_n + \sqrt{2}a_{N+1} \leq \sqrt{2}a_0 + 2560 \sum_{n=0}^N \frac{\mathcal{R}_z(s_n)}{s_n} \leq \sqrt{2} \operatorname{diam}(T) + 2560(4 + 2\sqrt{2})w_\mu(T).$$

Dropping the nonnegative term $\sqrt{2}a_{N+1}$ bounds every partial sum, and hence the full sum $\sum_{n \geq 0} a_n$, uniformly in z .

It remains to bound $\text{diam}(T)$. In the Gaussian case, the bound $\text{diam}(T) \leq \sqrt{2\pi} w(T)$ is standard; see [19, Proposition 7.5.2(f)]. For general μ , take $B = T$ and $n = 0$. The covering construction puts $\mathcal{G}_T$ in one ball of radius $\text{diam}(T)/4$. Since that ball has diameter $\text{diam}(T)/2$, we have $\mathcal{G}_T \neq T$. Choose $z \in T \setminus \mathcal{G}_T$. The definition of $\mathcal{G}_T$, monotonicity, and (1) give

$$\frac{\text{diam}(T)}{2560} \leq \frac{\mathcal{R}_z(5 \text{diam}(T))}{5 \text{diam}(T)} \leq \int_{5 \text{diam}(T)}^{\infty} \mathcal{R}_z(v) \frac{dv}{v^2} \leq 2w_\mu(T).$$

Hence $\text{diam}(T) \leq 5120w_\mu(T)$. Substituting this into the bound above and restoring the scaling proves Theorem 1, with $C = 10240(3 + 2\sqrt{2}) < 2^{16}$. $\blacksquare$

Remark 1. The construction follows the ideas of the fundamental partitioning theorem of Talagrand [4, Lemma 2.9.4 and Theorem 2.9.1]. In its Gaussian application, the points $z \in B$ to cover satisfy a condition of the form

$$\mathbb{E} \sup_{t \in B \cap (z + rB_2^d)} \langle G, t \rangle > \mathbb{E} \sup_{t \in B} \langle G, t \rangle - \eta, \quad (9)$$

where G is standard Gaussian and $r, \eta > 0$ are a prescribed radius and deficit. Informally, the neighborhood of z retains the expected supremum in the current cell up to the additive loss η . The geometric meaning of this criterion is less immediate, since it requires understanding Gaussian widths of localized subsets that change with the current cell. Our condition (5) asks whether perturbing z and projecting back onto the fixed set K returns close to z in mean square. Importantly, the profiles $s \mapsto \mathcal{R}_z(s)$ are fixed before constructing the partitions; the cell and its level determine only the scale and threshold.³ The projection integral bypasses an auxiliary growth functional and the selection of levels in [4, Lemma 2.9.5]. In the Gaussian case, KL testing gives the covering bound directly, without Sudakov minoration or Gaussian concentration.

Remark 2 (On probabilistic assumptions). For the lower bound, it suffices that one probability measure ρ and $\beta > 0$ satisfy

$$\mu_h(E) \leq \exp(\beta \|h\|^2) \rho(E) + \frac{1}{8} \quad (h \in \mathbb{R}^d, E \text{ Borel}). \quad (10)$$

After rescaling to $\beta = 1$, apply (10) to the disjoint events $(E_i - t_1)/s$ and shifts $(t_i - t_1)/s$, where $s = 5 \cdot 2^{-n/2} \text{diam}(B)$. The average success probability on disjoint events is at most $\frac{\exp(2^n/25)}{2^{2^n}} + \frac{1}{8} < \frac{7}{8}$. This gives $\gamma_2(T) \lesssim \sqrt{\beta} w_\mu(T)$. Condition (KL) implies (10) with $\rho = \mu$ and $\beta = 8\kappa$, by (4).

2 Subgaussian comparison from the projection integral

The classical subgaussian comparison theorem is a widely used consequence of the majorizing measures theorem, as in the original proof of Talagrand [11, Theorem 15]. The Gaussian comparison with arbitrary deterministic offsets already follows from Liu [7, Corollary 2]; van Handel [6] gives a streamlined exposition of Liu's proof mechanism and draws out its convex order consequences. Song and Talagrand [10] subsequently gave a very short argument deriving the convex order comparison from the seminorm specialization of the original majorizing measures theorem.

We give a short, self-contained proof of these comparison results, using neither the classical subgaussian comparison theorem nor results on stationary Gaussian processes. The reference

³Van Handel [15, §§3, 6.1] gives a simpler partition construction through a contraction principle. His Gaussian proof reorganizes the localized width argument using interpolation profiles defined through similar Gaussian width deficits; see (9) with $B = T$. These profiles are also fixed in advance. See also Remark 6.

experiment and calculations below are analogous to those in Pathak and Zhivotovskiy [3, proof of Lemma 3, (13) and (14)]. The proof also applies to general reference measures satisfying the translation bound (KL)⁴, with an explicit numerical constant.

Theorem 2. *Let $X \sim \mu$ be centered and integrable in $\mathbb{R}^d$, with $\mathcal{KL}(\mu_h \parallel \mu_k) \leq \kappa \|h - k\|^2$ for all $h, k \in \mathbb{R}^d$ and some $\kappa > 0$. Let $T \subset \mathbb{R}^d$ be finite and nonempty, and let $(U_t)_{t \in T}$ be a centered process such that, for some $L \geq 0$,*

$$\log \mathbb{E} \exp(\lambda(U_t - U_u)) \leq \frac{L^2 \lambda^2}{2} \|t - u\|^2 \quad (t, u \in T, \lambda \in \mathbb{R}). \quad (11)$$

Then

$$\mathbb{E} \max_{t \in T} U_t \leq 28L\sqrt{\kappa} w_\mu(T). \quad (12)$$

Taking $X = G$ and $\kappa = 1/2$ gives the classical Gaussian comparison.

Corollary 3. *Under the assumptions of Theorem 2, for every family of deterministic real numbers $(b_t)_{t \in T}$,*

$$\mathbb{E} \max_{t \in T} (U_t + b_t) \leq \mathbb{E} \max_{t \in T} \{28L\sqrt{\kappa} \langle X, t \rangle + b_t\}. \quad (13)$$

Proof of Theorem 2 and Corollary 3. If $L = 0$, all increments vanish almost surely and both claims are immediate. For $L > 0$, replace (U_t, b_t, X) by $(U_t/L, b_t/L, \sqrt{\kappa}X)$, so that $L = \kappa = 1$. It suffices to prove the corollary, since $b_t = 0$ gives the theorem. We use a fixed rule to break ties in all selectors below. Choose $Z \in \arg \max_{t \in T} (U_t + b_t)$, and take X independent of (U, Z) .

We first include the offsets in the projection identity (1). For $\sigma > 0$, set

$$\hat{t}_\sigma(y) \in \arg \min_{t \in T} \left\{ \|y - t\|^2 - \frac{2\sigma b_t}{28} \right\}, \quad \mathcal{R}(\sigma) = \mathbb{E} \|\hat{t}_\sigma(Z + \sigma X) - Z\|^2.$$

The same projection calculation as in (1) gives monotonicity of $\mathcal{R}$ and

$$\frac{1}{2} \int_0^\infty \mathcal{R}(\sigma) \frac{d\sigma}{\sigma^2} = \frac{1}{28} \left[\mathbb{E} \max_{t \in T} \{28\langle X, t \rangle + b_t\} - \mathbb{E} b_Z \right]. \quad (14)$$

Here $\mathbb{E}\langle X, Z \rangle = 0$ by independence and centering. It remains to prove $\mathbb{E} U_Z \leq 14 \int_0^\infty \mathcal{R}(\sigma) d\sigma / \sigma^2$.

We now compare two statistical experiments. Fix $t_0 \in T$ and positive scales $\sigma_0, \dots, \sigma_n$ with $\sigma_{k-1} = 2\sigma_k$. In the first experiment, take independent copies $X_1, \dots, X_n$ of X , independent of U , and set

$$Y_k = Z + \sigma_k X_k, \quad t_k = \hat{t}_{\sigma_k}(Y_k) \quad (1 \leq k \leq n).$$

Then $\mathbb{E} \|t_k - Z\|^2 = \mathcal{R}(\sigma_k)$. Let P_k denote the joint law of the process U and the first k observations $(Y_1, \dots, Y_k)$ in this experiment.

We define the reference law Q_k on the same variables $(U, Y_1, \dots, Y_k)$ by keeping the law of U and generating observations successively with

$$\mathcal{L}_{Q_k}(Y_j \mid U, Y_1, \dots, Y_{j-1}) = \mathcal{L}(t_{j-1} + \sigma_j X), \quad 1 \leq j \leq k,$$

using fresh noise at each step. Here t_0 is fixed and $t_j = \hat{t}_{\sigma_j}(Y_j)$ is the same deterministic function of the observation under either law. Under Q_k , this recursion never uses U , so all observations and

⁴The Gaussian comparison is the main case of interest. Allowing other reference laws may also be useful when studying the best numerical constants, through the quantity $\sqrt{\kappa} w_\mu(T)$.

selected points are independent of U . Under P_k , they can depend on U through Z . At level 0, put $P_0 = Q_0 = \mathcal{L}(U)$.

Since the law of U is the same under both experiments, the likelihood ratio factors over the conditional laws of the observations:

$$\log \frac{dP_k}{dQ_k}(U, Y_1, \dots, Y_k) = \sum_{j=1}^k \log \frac{d\mu_{Z/\sigma_j}}{d\mu_{t_{j-1}/\sigma_j}}(Y_j/\sigma_j).$$

Under P_k , conditionally on $U, Y_1, \dots, Y_{j-1}$, the variable Y_j/σ_j has law μ_{Z/σ_j} . Taking conditional expectations and applying (KL), we obtain

$$\mathcal{KL}(P_k \| Q_k) - \mathcal{KL}(P_{k-1} \| Q_{k-1}) = \mathbb{E} \mathcal{KL}(\mu_{Z/\sigma_k} \| \mu_{t_{k-1}/\sigma_k}) \leq \frac{\mathbb{E} \|t_{k-1} - Z\|^2}{\sigma_k^2}. \quad (15)$$

Independence under Q_k and (11) give

$$\mathbb{E}_{Q_k} \exp \left(\frac{U_{t_k} - U_{t_{k-1}}}{\sigma_k} - \frac{\|t_k - t_{k-1}\|^2}{2\sigma_k^2} \right) \leq 1.$$

The change of measure inequality from the proof of Lemma 1 bounds the P_k expectation of the exponent by $\mathcal{KL}(P_k \| Q_k)$. Multiplying by σ_k , adding $\sigma_k \mathcal{KL}(P_k \| Q_k)$, and using $\sigma_{k-1} = 2\sigma_k$ gives

$$\begin{aligned} \mathbb{E} U_{t_k} + \sigma_k \mathcal{KL}(P_k \| Q_k) &\leq \mathbb{E} U_{t_{k-1}} + \sigma_{k-1} \mathcal{KL}(P_{k-1} \| Q_{k-1}) + \frac{\mathbb{E} \|t_k - t_{k-1}\|^2}{2\sigma_k} \\ &\quad + 2\sigma_k [\mathcal{KL}(P_k \| Q_k) - \mathcal{KL}(P_{k-1} \| Q_{k-1})] \\ &\leq \mathbb{E} U_{t_{k-1}} + \sigma_{k-1} \mathcal{KL}(P_{k-1} \| Q_{k-1}) + \frac{\mathcal{R}(\sigma_k) + 3\mathbb{E} \|t_{k-1} - Z\|^2}{\sigma_k}. \end{aligned}$$

The last step uses (15) and $\|t_k - t_{k-1}\|^2 \leq 2\|t_k - Z\|^2 + 2\|t_{k-1} - Z\|^2$. We now sum over k . The left and right sides telescope; using $P_0 = Q_0$, $\mathbb{E} U_{t_0} = 0$, and $\sigma_n \mathcal{KL}(P_n \| Q_n) \geq 0$, we obtain

$$\begin{aligned} \mathbb{E} U_{t_n} &\leq \sum_{k=1}^n \frac{\mathcal{R}(\sigma_k) + 3\mathbb{E} \|t_{k-1} - Z\|^2}{\sigma_k} \\ &= \frac{6\mathbb{E} \|t_0 - Z\|^2}{\sigma_0} + 7 \sum_{k=1}^{n-1} \frac{\mathcal{R}(\sigma_k)}{\sigma_k} + \frac{\mathcal{R}(\sigma_n)}{\sigma_n} \\ &\leq \frac{6 \text{diam}(T)^2}{\sigma_0} + 7 \sum_{k=1}^n \frac{\mathcal{R}(\sigma_k)}{\sigma_k}. \end{aligned} \quad (16)$$

The equality uses $\sigma_{k-1} = 2\sigma_k$ and the definition of $\mathcal{R}$. For $m \geq 1$, take $n = 2m$ and $\sigma_k = 2^{m-k}$ in (16). For a fixed copy of X independent of U , finiteness of T gives $\hat{t}_{2^{-m}}(Z + 2^{-m}X) = Z$ for all sufficiently large m , almost surely. Therefore, as $m \rightarrow \infty$,

$$\mathbb{E} U_{t_n} = \mathbb{E} U_{\hat{t}_{2^{-m}}(Z + 2^{-m}X)} \longrightarrow \mathbb{E} U_Z.$$

Finally, monotonicity gives $\mathcal{R}(s)/s \leq 2 \int_s^{2s} \mathcal{R}(\sigma) d\sigma/\sigma^2$. Letting $m \rightarrow \infty$ therefore yields

$$\mathbb{E} U_Z \leq 7 \sum_{k \in \mathbb{Z}} \frac{\mathcal{R}(2^{-k})}{2^{-k}} \leq 14 \int_0^\infty \mathcal{R}(\sigma) \frac{d\sigma}{\sigma^2} = \mathbb{E} \max_{t \in T} \{28 \langle X, t \rangle + b_t\} - \mathbb{E} b_Z.$$

Since $U_Z + b_Z = \max_{t \in T} (U_t + b_t)$, adding $\mathbb{E} b_Z$ and undoing the normalization proves the corollary. $\blacksquare$

Remark 3. For any centered vector Y satisfying $\log \mathbb{E} \exp(\langle v, Y \rangle) \leq L^2 \|v\|^2/2$ for every $v \in \mathbb{R}^d$, Corollary 3 gives, using the standard approximation argument as in [6, §1.1],

$$\mathbb{E} f(Y) \leq \mathbb{E} f(28L\sqrt{\kappa} X) \quad \text{for every convex } f : \mathbb{R}^d \rightarrow \mathbb{R}.$$

3 A geometric construction of admissible sequences

We now give a geometric proof of convex hull stability of γ_2 for finite subsets of a Hilbert space, addressing one possible interpretation of the question of Talagrand [4, Research Problem 2.11.2]. We essentially adapt Section 1 to uniform measure on a Euclidean ball. The argument still passes through an averaged linear supremum, but testing becomes a volume comparison and metric projections define the split.

Remark 4. Talagrand suggests that a more satisfactory geometric argument would extend convex hull stability to 2-smooth Banach spaces [4, Research Problem 2.11.3]. Zhang and Zhao [20] show, however, that this extension fails in $L^p[0, 1]$ and ℓ_p for every $2 < p < \infty$.

The input admissible sequence is used only to bound the diameter and ball averaged supremum of $K = \text{conv } T$. We then construct partitions from K itself: in each cell, we cover points whose perturbations usually project back close to them, and keep the uncovered remainder as one child. The projection integral controls the total remainder cost.

Theorem 4. *Given an admissible sequence of a nonempty finite subset T of a Hilbert space with cost $L < \infty$, the construction below produces an admissible sequence of $K = \text{conv } T$ with cost at most $2^{16}L$. Consequently*

$$\gamma_2(T) \leq \gamma_2(\text{conv } T) \leq 2^{16}\gamma_2(T).$$

The proof uses only inner products, norms, convex hulls and metric projections; translations, dilations, intersections and Euclidean volume comparison; counting, averaging, elementary inequalities and calculus in one variable. Probabilities and expectations are used only as shorthand for volume ratios and averages.

Proof. We first use the input admissible sequence to bound the linear supremum averaged over the Euclidean ball. We then use the projection integral and volume coverings to construct an admissible sequence for K with cost at most $2^{16}L$, following the splitting rule of Section 1. The singleton case is immediate. Translate a point of T to 0 and identify its span with $\mathbb{R}^d$. Throughout this section, X is uniformly distributed on $\sqrt{d}B_2^d$ in the Euclidean space under consideration. We write $|\cdot|$ for Euclidean volume in the appropriate dimension.

Write $(\mathcal{A}_n)$ for the given admissible sequence for T . For each cell $F \in \mathcal{A}_n$, choose a representative $r_F \in F$, taking $r_T = 0$ for the root cell at level 0. For $n \geq 1$, let F^- be the unique cell of $\mathcal{A}_{n-1}$ containing F ; this is its parent. Define

$$v_F = r_F - r_{F^-}.$$

Both representatives lie in F^- , so $\|v_F\| \leq \text{diam}(F^-)$. For each $t \in T$, let $A_n(t)$ be its cell in $\mathcal{A}_n$. Since the given sequence has cost L , we have $\|t - r_{A_n(t)}\| \leq \text{diam}(A_n(t)) \leq 2^{-n/2}L \rightarrow 0$. Using $r_T = 0$ gives

$$t = \sum_{n \geq 1} v_{A_n(t)}, \quad \sum_{n \geq 1} 2^{n/2} \text{diam}(A_{n-1}(t)) \leq \sqrt{2}L.$$

For a unit vector e and $0 \leq v \leq \sqrt{d}$, consider the cap $\{x \in \sqrt{d}B_2^d : \langle x, e \rangle > v\}$. Since

$$\|x - ve\|^2 = \|x\|^2 - 2v\langle x, e \rangle + v^2 \leq d - v^2,$$

the translation $x \mapsto x - ve$ maps this cap into the half-ball $\{y \in \sqrt{d - v^2}B_2^d : \langle y, e \rangle > 0\}$, because $\langle x - ve, e \rangle > 0$. Translation preserves volume, so the cap has at most half the volume of $\sqrt{d - v^2}B_2^d$.

The opposite cap has the same volume by reflection. Their combined volume is therefore at most $|\sqrt{d} - v^2 B_2^d|$. Dividing by $|\sqrt{d} B_2^d|$ gives

$$\mathbb{P}(|\langle X, e \rangle| > v) \leq (1 - v^2/d)^{d/2} \leq \exp(-v^2/2). \quad (17)$$

For $v > \sqrt{d}$ the caps are empty. By the decomposition of t above, if $\max_{t \in T} |\langle x, t \rangle| > \sqrt{2}L(2+u)$, then some level $n \geq 1$ and cell $F \in \mathcal{A}_n$ satisfy $|\langle x, v_F \rangle| > (2+u)2^{n/2} \text{diam}(F^-)$. At level n there are at most 2^{2^n} such vectors, each of norm at most $\text{diam}(F^-)$. The cap estimate and $(2+u)^2/2 - \log 2 \geq 1+u$ therefore give, for $u \geq 0$,

$$\mathbb{P}\left(\max_{t \in T} |\langle X, t \rangle| > \sqrt{2}L(2+u)\right) \leq \sum_{n \geq 1} \exp(-(1+u)2^n) \leq \exp(-u).$$

The last inequality follows from $2^n \geq n+1$ and summing a geometric series. Convexity enters through the pointwise identity

$$\max_{t \in K} \langle x, t \rangle = \max_{t \in T} \langle x, t \rangle. \quad (18)$$

Since $0 \in T$, the supremum is nonnegative. Integrating the preceding bound therefore gives

$$\mathbb{E} \max_{t \in K} \langle X, t \rangle \leq \sqrt{2}L \left(2 + \int_0^\infty \exp(-u) du\right) = 3\sqrt{2}L < 5L. \quad (19)$$

The same identity gives $\text{diam}(K) = \text{diam}(T) \leq L$, since the diameter of a set is the supremum, over unit directions, of its linear oscillation. The input sequence is used only to establish these two bounds. The splitting rule below depends on K , the current cell, and its level, but not on the input partitions or L .

The averaged projection integral. For $z \in K$ and $s > 0$, put

$$\mathcal{R}_z(s) = \mathbb{E} \|\Pi_K(z + sX) - z\|^2.$$

Thus $\mathcal{R}_z(s)$ is the mean squared projection error for points uniformly distributed over $z + s\sqrt{d} B_2^d$, measured from z . We take expectations in the pointwise projection decomposition [1, equation (15b)]. Since $\mathbb{E}X = 0$ by symmetry, $\mathbb{E}\langle X, z \rangle = 0$, giving for every $z \in K$

$$\frac{1}{2} \int_0^\infty \mathcal{R}_z(v) \frac{dv}{v^2} = \mathbb{E} \max_{t \in K} \langle X, t \rangle < 5L. \quad (20)$$

For positive scales with $s_{j+1} \leq s_j/\sqrt{2}$, apply the same calculation over disjoint intervals as in (2). Since $\sqrt{2}/(\sqrt{2}-1) < 7/2$, we obtain

$$\sum_j \frac{\mathcal{R}_z(s_j)}{s_j} \leq \frac{7}{2} \int_0^\infty \mathcal{R}_z(v) \frac{dv}{v^2} \leq 35L. \quad (21)$$

Covering points with small projection error. We repeat the testing step of Section 1: separated points with small projection error give disjoint success sets. Here a volume comparison bounds their number. Fix a set $B \subset K$ with $\text{diam}(B) > 0$ and an integer $n \geq 2$. Let $s_n = 2 \text{diam}(B) 2^{-n/2}$.

We now introduce one of our key geometric objects. We call $z \in B$ *good* if

$$\mathbb{P}\left(\|\Pi_K(z + s_n X) - z\| \leq \frac{\text{diam}(B)}{8}\right) > \frac{3}{4}. \quad (22)$$

In words, a point is *good* if projecting $z + s_n X$ back onto K returns a point within $\text{diam}(B)/8$ of z with probability greater than $3/4$.

We claim that these points are covered by at most $2^{2^n} - 1$ closed balls of radius $\text{diam}(B)/4$. Otherwise, successive selection gives 2^{2^n} such points $t_1, \dots, t_{2^{2^n}}$ at pairwise distance greater than $\text{diam}(B)/4$. Put

$$\mathcal{D}_n = t_1 + \sqrt{ds_n^2 + \text{diam}(B)^2} B_2^d.$$

For $x \in \sqrt{d} B_2^d$, the parallelogram identity gives

$$\|t_i + s_n x - t_1\|^2 + \|t_i - s_n x - t_1\|^2 = 2\|t_i - t_1\|^2 + 2s_n^2 \|x\|^2 \leq 2(ds_n^2 + \text{diam}(B)^2).$$

At least one of $t_i + s_n x$ and $t_i - s_n x$ therefore lies in $\mathcal{D}_n$. By symmetry, $\mathbb{P}(t_i + s_n X \in \mathcal{D}_n) \geq 1/2$. Together with condition (22), this shows that the sets

$$E_i = \mathcal{D}_n \cap (t_i + s_n \sqrt{d} B_2^d) \cap \{y : \|\Pi_K(y) - t_i\| \leq \text{diam}(B)/8\}$$

satisfy $\mathbb{P}(t_i + s_n X \in E_i) > 1/4$, hence $|E_i| > s_n^d |\sqrt{d} B_2^d|/4$. They are pairwise disjoint. Since all E_i lie in $\mathcal{D}_n$, summing their volumes gives

$$2^{2^n} < \frac{4|\mathcal{D}_n|}{s_n^d |\sqrt{d} B_2^d|} = 4 \left(1 + \frac{\text{diam}(B)^2}{ds_n^2}\right)^{d/2} \leq 4 \exp(\text{diam}(B)^2/(2s_n^2)) = 4 \exp(2^n/8) < 2^{2^n}. \quad (23)$$

The last inequality holds for $n \geq 2$, giving a contradiction.

If z is not *good*, the projection error exceeds $\text{diam}(B)/8$ with probability at least $1/4$. Consequently,

$$\mathcal{R}_z(s_n) \geq \frac{1}{4} \left(\frac{\text{diam}(B)}{8}\right)^2 = \frac{\text{diam}(B)^2}{256},$$

and hence

$$2^{n/2} \text{diam}(B) = \frac{2 \text{diam}(B)^2}{s_n} \leq 512 \frac{\mathcal{R}_z(s_n)}{s_n}. \quad (24)$$

The construction of admissible partitions. Start with $\mathcal{B}_0 = \mathcal{B}_1 = \mathcal{B}_2 = \{K\}$. Given $\mathcal{B}_n$ for $n \geq 2$, take a cell $B \in \mathcal{B}_n$ of positive diameter and set $s_n = 2 \text{diam}(B) 2^{-n/2}$. Choose centers c_i satisfying the projection condition (22), each outside the preceding closed balls $C_j = c_j + (\text{diam}(B)/4) B_2^d$, until all such points are covered. The covering argument forces this procedure to stop with $\ell \leq 2^{2^n} - 1$ centers. As in Section 1, replace B by the nonempty sets among

$$(B \cap C_i) \setminus \bigcup_{j < i} C_j \quad (1 \leq i \leq \ell), \quad B \setminus \bigcup_{i=1}^{\ell} C_i.$$

The same counting argument as in (6) gives refining admissible partitions: for $N > 2$,

$$|\mathcal{B}_N| \leq \prod_{n=2}^{N-1} 2^{2^n} = 2^{2^N - 4} \leq 2^{2^N}.$$

Finally, fix $z \in K$, let $B_n(z)$ be its cell in $\mathcal{B}_n$, and put

$$a_n = 2^{n/2} \text{diam}(B_n(z)), \quad s_n = 2 \cdot 2^{-n/2} \text{diam}(B_n(z)).$$

As in (8), a covered child has at most half the parent diameter, while the remainder is controlled by (24). Thus, at every level $n \geq 2$ of positive diameter,

$$\sqrt{2}a_{n+1} \leq a_n + 512 \frac{\mathcal{R}_z(s_n)}{s_n}.$$

The scales satisfy $s_{n+1} \leq s_n/\sqrt{2}$. The input bound (19), through the projection identity (20) and (21), therefore bounds $\sum_{n \geq 2: s_n > 0} \mathcal{R}_z(s_n)/s_n$ by $35L$. The remaining summation is the same as in Section 1: for $N \geq 2$ with $a_N > 0$, summing over $2 \leq n \leq N$ and using $a_0 = \text{diam}(K)$, $a_1 = \sqrt{2} \text{diam}(K)$, and $a_2 = 2 \text{diam}(K)$ gives

$$\begin{aligned} (\sqrt{2} - 1) \sum_{n=0}^N a_n + \sqrt{2}a_{N+1} &\leq (1 + 2\sqrt{2}) \text{diam}(K) + 512 \sum_{n=2}^N \frac{\mathcal{R}_z(s_n)}{s_n} \\ &\leq (1 + 2\sqrt{2}) \text{diam}(K) + 17920L < (\sqrt{2} - 1)2^{16}L. \end{aligned}$$

Every partial sum, and hence the full sum, is bounded by $2^{16}L$, uniformly in z . This proves the claimed cost bound. Restricting partitions of K to T gives $\gamma_2(T) \leq \gamma_2(K)$. $\blacksquare$

The construction in the proof of Theorem 4 does not require convexity or an input admissible sequence. The input sequence above is used only to bound its cost. We now state the general form of the construction; the sets to be covered can be defined by the projection condition (22), or by the finite average in Remark 5 below.

Corollary 5. *Let $H \subset \mathbb{R}^d$ be nonempty and compact, and fix a measurable nearest point selector Π_H . The projection and covering rule using (22), with H in place of K , constructs an admissible sequence of cost at most $2^{16}\gamma_2(H)$.*

Proof. As noted after (1), the projection identity and monotonicity hold for non-convex H . The volume argument (23) only requires a fixed selector, so that the success sets of separated points are disjoint. Thus the covering rule and the branch summation apply unchanged. Apply (19) to an arbitrary admissible sequence of H and take the infimum over its cost. $\blacksquare$

Remark 5 (Admissible sequences from finite projections). The sets to be covered can also be defined by a finite average of squared projection distances, even for nonconvex H . Choose once, depending only on d , a symmetric $1/64$ -net $x_1, \dots, x_N$ of $\sqrt{d}B_2^d$, with $N \leq (1 + 128\sqrt{d})^d$. Fix w_j as the probability that x_j is a closest net point to X :

$$w_j = \mathbb{P}(\|X - x_j\| \leq \|X - x_i\| \text{ for all } 1 \leq i \leq N).$$

In a cell B of positive diameter at level $n \geq 2$, put $s_n = 2 \text{diam}(B)2^{-n/2}$ and cover the set

$$\mathcal{G}_B = \left\{ z \in B : \sum_{j=1}^N w_j \|\Pi_H(z + 2s_n x_j) - z\|^2 < \frac{\text{diam}(B)^2}{1024} \right\}, \quad (25)$$

retaining the uncovered points as before. The resulting admissible sequence has cost at most $2^{19}\gamma_2(H)$. Once the points and weights are fixed, the recursive construction requires no volume evaluations: only nearest point projections onto H , finite weighted sums, and successive ball coverings. With this definition, the rest of the proof proceeds as above; we omit the details.

4 Examples

We now turn to several related questions raised in [4]. The next two examples give geometric proofs in the sense of Section 3. Research Problem 2.11.8 asks for a geometric proof of the γ_2 bound for convex hulls of vectors with logarithmically decreasing norms.

Example 6. Let $(t_j)_{j \geq 1}$ be Hilbert vectors with $\|t_j\| \leq A/\sqrt{\log(j+1)}$. Then

$$\gamma_2(\text{conv}\{\pm t_j : j \geq 1\}) \leq 2^{16}A.$$

Proof. To apply the construction, take a finite hull $K = \text{conv}\{\pm t_1, \dots, \pm t_N\}$ in its span $\mathbb{R}^d$; the zero case is immediate. Let X be uniform on $\sqrt{d}B_2^d$. The estimate (17) and the norm bound give, for $u \geq 0$,

$$\mathbb{P}\left(\max_{t \in K} \langle X, t \rangle > (2+u)A\right) \leq \sum_{j \geq 1} (j+1)^{-2-2u} \leq \exp(-u).$$

For the last inequality, we use $(j+1)^{-2u} \leq 2^{-2u} \leq \exp(-u)$ and $\sum_{j \geq 1} (j+1)^{-2} < 1$. Integrating gives

$$\mathbb{E} \max_{t \in K} \langle X, t \rangle \leq 3A.$$

Insert this bound and $\text{diam}(K) \leq 2A/\sqrt{\log 2}$ into the branch estimate underlying Corollary 5 to obtain $2^{16}A$. The admissible partition construction uses condition (22). ■

Research Problem 2.11.13 asks for a geometric proof of subadditivity of γ_2 under Minkowski addition. Research Problem 2.11.14 asks for the special bound $\gamma_2(\sum_{i=1}^N T_i) \leq CN\sqrt{\log M}$ when each T_i consists of M unit vectors. The next example gives both. Only additivity of linear suprema is needed; the covering argument is unchanged.

Example 7. Let $T_1, \dots, T_N$ be nonempty finite subsets of a common Hilbert space, with given admissible sequences of finite costs $L_1, \dots, L_N$. The construction above produces an admissible sequence of $\sum_{i=1}^N \text{conv} T_i$ with cost at most $2^{16} \sum_{i=1}^N L_i$. Consequently, for arbitrary bounded nonempty Hilbert sets T_i ,

$$\gamma_2\left(\sum_{i=1}^N T_i\right) \leq 2^{16} \sum_{i=1}^N \gamma_2(T_i), \quad \sum_{i=1}^N T_i = \{t_1 + \dots + t_N : t_i \in T_i\}.$$

Proof. Translate one point of each T_i to 0 and work in their common span $\mathbb{R}^d$. Put $K = \sum_i \text{conv} T_i$. Let X be uniform on $\sqrt{d}B_2^d$. Since

$$\max_{t \in K} \langle x, t \rangle = \sum_i \max_{t \in T_i} \langle x, t \rangle,$$

applying (19) to each input sequence gives

$$\text{diam}(K) \leq \sum_i L_i, \quad \mathbb{E} \max_{t \in K} \langle X, t \rangle \leq 3\sqrt{2} \sum_i L_i < 5 \sum_i L_i.$$

The construction therefore applies with $L = \sum_i L_i$. Restrict the resulting partitions to $\sum_i T_i$ and take the infimum over the input costs; compactness gives the general case. For Research Problem 2.11.14, keep each T_i in one cell until its singleton partition satisfies the admissibility bound. Since $\text{diam}(T_i) \leq 2$, the geometric sum of the preceding levels gives cost at most $C\sqrt{\log M}$ for $M \geq 2$. The construction above then gives cost at most $CN\sqrt{\log M}$; for $M = 1$, the sum is a singleton. ■

We now apply the same projection and covering rule directly to a union of ellipsoids. Let T be a nonempty finite set of real $m \times d$ matrices, and put

$$d_\infty(t, r) = \|t - r\|_{\text{op}}, \quad E_t = t^\top B_2^m, \quad H = \bigcup_{t \in T} E_t.$$

Here $t^\top$ is the transpose of t , so $E_t \subset \mathbb{R}^d$ is the ellipsoid obtained by applying $t^\top$ to the Euclidean unit ball. The Hilbert–Schmidt norm $\|t\|_{\text{HS}}$ is the square root of the sum of squared matrix entries, and $\|t\|_{\text{op}}$ is the Euclidean operator norm. A consequence of the chaos bounds of Krahmer, Mendelson, and Rauhut [13], recorded in [4, Proposition 15.1.21], is

$$\left(\mathbb{E} \sup_{t \in T} \|tG\|^2 \right)^{1/2} \leq C \left(\sup_{t \in T} \|t\|_{\text{HS}} + \gamma_2(T, d_\infty) \right). \quad (26)$$

Since $\sup_t \|tG\| = \sup_{h \in H} \langle G, h \rangle$, Talagrand [4, Research Problem 15.1.22] asks for a direct construction of an admissible sequence on H whose cost is bounded by the right side of (26). We construct the partitions geometrically and, for brevity, use the cited bound (26) to estimate their cost.

Example 8. The projection and covering construction produces an admissible sequence of H with cost at most

$$C \left(\sup_{t \in T} \|t\|_{\text{HS}} + \gamma_2(T, d_\infty) \right),$$

where C is universal.

Proof. We apply Corollary 5, using the nearest point selector described below.

Projection onto the union. For $t \in T$, let $b_i > 0$ be its positive singular values and v_i the corresponding right singular vectors. In these coordinates,

$$E_t = \left\{ \sum_i a_i v_i : \sum_i \frac{a_i^2}{b_i^2} \leq 1 \right\}.$$

Write $y_i = \langle y, v_i \rangle$. If $\sum_i y_i^2 / b_i^2 \leq 1$, then $\Pi_{E_t}(y) = \sum_i y_i v_i$. Otherwise,

$$\Pi_{E_t}(y) = \sum_i \frac{b_i^2}{b_i^2 + \lambda} y_i v_i, \quad \sum_i \frac{b_i^2 y_i^2}{(b_i^2 + \lambda)^2} = 1, \quad \lambda > 0. \quad (27)$$

We now define a nearest point selector by

$$\Pi_H(y) \in \arg \min_{q \in H} \|y - q\|^2 = \arg \min_{q \in \{\Pi_{E_t}(y) : t \in T\}} \|y - q\|^2.$$

Projection ties are resolved by a fixed ordering of the matrices. Evaluating Π_H therefore requires only the scalar equations (27) and comparison of the resulting distances.

The partition rule. Use this selector in the projection condition (22) or the finite test (25). In each cell B , cover the selected points by balls of radius $\text{diam}(B)/4$ and retain the uncovered remainder as one child. The preceding results give admissibility and the cost bound; only the projection formulas are specific to the ellipsoids.

The cost. Let X be uniform on $\sqrt{d} B_2^d$. Since $\max_{h \in H} \langle x, h \rangle = \max_{t \in T} \|tx\|$, the branch estimate used in Corollary 5 and its finite version bound the cost by

$$C \left(\text{diam}(H) + \mathbb{E} \max_{t \in T} \|tX\| \right).$$

It remains to insert the known bound. Cauchy–Schwarz and radial integration give

$$\mathbb{E} \max_t \|tX\| \leq \left(\mathbb{E} \max_t \|tX\|^2 \right)^{1/2} = \sqrt{\frac{d}{d+2}} \left(\mathbb{E} \max_t \|tG\|^2 \right)^{1/2}.$$

For the equality, use homogeneity and the mean squared radii $\mathbb{E}\|X\|^2 = d^2/(d+2)$ and $\mathbb{E}\|G\|^2 = d$. Together with $\text{diam}(H) \leq 2 \sup_t \|t\|_{\text{HS}}$ and (26), this proves the asserted cost. $\blacksquare$

4.1 Special geometries where good sets are explicit

The constructions in Section 3 are geometric, but in special cases we can make the *good* sets more explicit. Talagrand [4, §2.14] gives an explicit chaining construction of optimal order tailored to ellipsoids. Here we apply the splitting rule of Section 3 to ellipsoids and ℓ_1 balls, replacing the projection conditions (5), (22), or (25) by explicit coordinate conditions. The sets depend only on the current cell and level, and their covering bounds follow from coordinate packing.

Example 9. Let

$$E = \left\{ z \in \mathbb{R}^d : \sum_{i=1}^d \frac{z_i^2}{a_i^2} \leq 1 \right\}, \quad a_i > 0, \quad \|a\| = \left(\sum_{i=1}^d a_i^2 \right)^{1/2}.$$

In a cell $B \subset E$ with $\text{diam}(B) > 0$ at level $n \geq 2$, put $s_n = 2 \text{diam}(B) 2^{-n/2}$ and cover the *good* set

$$\mathcal{G}_B = \left\{ z \in B : \sum_{a_i^2 < \|a\| s_n} z_i^2 + \sum_{i=1}^d \min \left\{ s_n^2, \frac{a_i^4}{\|a\|^2} \right\} < \frac{\text{diam}(B)^2}{256} \right\}. \quad (28)$$

The splitting rule of Section 3, starting with $\mathcal{B}_0 = \mathcal{B}_1 = \mathcal{B}_2 = \{E\}$, constructs admissible partitions of cost at most $C\|a\|$, which is of optimal order.

Proof. We first check the covering bound. Put $J = \{i : a_i^2 \geq \|a\| s_n\}$. If $\mathcal{G}_B$ is nonempty, then $|J| s_n^2 < \text{diam}(B)^2/256$, so $|J| < 2^n/1024$, and every $z \in \mathcal{G}_B$ satisfies $\|z_{J^c}\| < \text{diam}(B)/16$. For any two points z, z' in a $\text{diam}(B)/4$ -separated family of N points in $\mathcal{G}_B$, the triangle inequality gives

$$\|(z - z')_J\| \geq \|z - z'\| - \|(z - z')_{J^c}\| > \frac{\text{diam}(B)}{4} - \frac{\text{diam}(B)}{8} = \frac{\text{diam}(B)}{8}.$$

For $J \neq \emptyset$, disjoint balls of radius $\text{diam}(B)/16$ around the projected centers lie in a ball of radius $17 \text{diam}(B)/16$ around one of these centers, giving

$$N \leq 17^{|J|} < 2^{5|J|} < 2^{2^n}.$$

For $J = \emptyset$, one center suffices. Thus successively choosing uncovered points of $\mathcal{G}_B$ gives a cover by at most $2^{2^n} - 1$ balls of radius $\text{diam}(B)/4$.

We now put

$$U_z(s) = \sum_{a_i^2 < \|a\| s} z_i^2 + \sum_{i=1}^d \min \left\{ s^2, \frac{a_i^4}{\|a\|^2} \right\}.$$

This is the left side of the inequality in (28); it replaces $\mathcal{R}_z$ in the branch estimate. Each term is nondecreasing in s . Integrating the first sum over $s > a_i^2/\|a\|$ and splitting each term of the second at $s = a_i^2/\|a\|$ gives the following counterpart of the projection integral for ellipsoids:

$$\int_0^\infty U_z(s) \frac{ds}{s^2} = \sum_{i=1}^d \left(\frac{\|a\| z_i^2}{a_i^2} + \frac{2a_i^2}{\|a\|} \right) = \|a\| \left(2 + \sum_{i=1}^d \frac{z_i^2}{a_i^2} \right) \leq 3\|a\|. \quad (29)$$

Along a branch, $s_{n+1} \leq s_n/\sqrt{2}$, so the disjoint-interval calculation in (2) gives $\sum_n U_z(s_n)/s_n \leq 21\|a\|/2$. If z belongs to the remainder child, then

$$2^{n/2} \text{diam}(B) = \frac{2 \text{diam}(B)^2}{s_n} \leq 512 \frac{U_z(s_n)}{s_n}.$$

A covered child has diameter at most $\text{diam}(B)/2$. The summation in Section 3, with U_z in place of $\mathcal{R}_z$ and $\text{diam}(E) \leq 2\|a\|$, now gives cost at most $C\|a\|$. For optimality, see [4, §2.14]. ■

Remark 6 (Comparison with the interpolation approach). For a symmetric convex body K , write $\|y\|_K = \inf\{r > 0 : y \in rK\}$ for its gauge. Van Handel [14, Propositions 2.3 and 2.5] constructs nets by covering the images $K_s = \pi_s(K)$ of interpolation minimizers $\pi_s(z) \in \arg \min_y \{\|y\|_K + s\|z - y\|\}$ and integrating their approximation errors. The bound [14, Theorem 1.2] is, for every $a > 0$,

$$\gamma_2(K) \leq C \left(\frac{1}{a} + \sum_{n \geq 0} 2^{n/2} e_n(K_{a2^{n/2}}) \right).$$

Here $e_n(A)$ is the infimum radius of a cover of A by fewer than 2^{2^n} balls.

For ellipsoids, the specific geometry of the interpolation sets yields the sharp bound [14, Proposition 3.1]. However, the interpolation entropy bound is not sharp in general, even in finite dimensions. For K_ε in Example 10 and fixed $0 < \varepsilon < 1$, the displayed bound gives order $(\log d)^{3/2}$ instead of $\sqrt{\log d}$; see [14, §3.3]. Example 10 below gives a geometric proof and a direct construction of admissible partitions using the coordinate condition (30), which replaces the general condition (22) for this example. The approach of Section 3 gives cost at most $C\gamma_2(K)$ for every compact $K \subset \mathbb{R}^d$.

Example 10. Let $0 \leq \varepsilon \leq 1$, put $u = d^{-1/2}(1, \dots, 1)$, and set

$$K_\varepsilon = \text{conv}\{\pm(e_i + \varepsilon u) : 1 \leq i \leq d\}, \quad \lambda = \sqrt{\log(ed)}.$$

The coordinates of $z \in K_\varepsilon$ in the basis $(e_i + \varepsilon u)_{i=1}^d$ are

$$a_i(z) = z_i - \frac{\varepsilon \langle z, u \rangle}{1 + \varepsilon \sqrt{d}}, \quad 1 \leq i \leq d.$$

Hence $z = \sum_i a_i(z)(e_i + \varepsilon u)$ and $\|a(z)\|_1 \leq 1$. In a cell $B \subset K_\varepsilon$ with $\text{diam}(B) > 0$ at level $n \geq 2$, put $s_n = 2 \text{diam}(B) 2^{-n/2}$ and cover the *good* set

$$\mathcal{G}_B = \left\{ z \in B : \sum_{i=1}^d \min\{a_i(z)^2, \lambda^2 s_n^2\} < \frac{\text{diam}(B)^2}{256} \right\}. \quad (30)$$

The splitting rule, starting with $\mathcal{B}_0 = \mathcal{B}_1 = \mathcal{B}_2 = \{K_\varepsilon\}$, constructs admissible partitions of cost at most $C\sqrt{\log(ed)}$, of optimal order uniformly in ε .

Proof. We first check the covering bound. For $z \in \mathcal{G}_B$, put $I = \{i : |a_i(z)| \geq \lambda s_n\}$. Then

$$|I| \lambda^2 < \frac{2^n}{1024}, \quad \|a(z)_{I^c}\| < \frac{\text{diam}(B)}{16}.$$

Let m be the largest integer in $\{0, \dots, d\}$ satisfying $m\lambda^2 < 2^n/1024$, and enlarge each I to a set J of size m . The distance from z to $S_J = \text{span}(u, \{e_i : i \in J\})$ is less than $\text{diam}(B)/16$: indeed, $a(z)_J + \varepsilon(\sum_i a_i(z))u \in S_J$. Within a $\text{diam}(B)/4$ -separated family in $\mathcal{G}_B$, the points assigned

to the same J have orthogonal projections onto S_J separated by more than $\text{diam}(B)/5$, since $(\text{diam}(B)/4)^2 - (\text{diam}(B)/8)^2 > (\text{diam}(B)/5)^2$. These projections have diameter at most $\text{diam}(B)$, so disjoint balls of radius $\text{diam}(B)/10$ give at most 11^{m+1} points for each J . The total number is at most

$$\binom{d}{m} 11^{m+1} \leq 11 \exp(m[\log(ed) + \log 11]) \leq 11 \exp((1 + \log 11)2^n/1024) < 2^{2^n},$$

where the last inequality holds for $n \geq 2$. Thus at most $2^{2^n} - 1$ balls of radius $\text{diam}(B)/4$ cover $\mathcal{G}_B$. Keeping the direction u in every S_J avoids any loss from the stretching along u .

We now put

$$U_z(s) = \sum_{i=1}^d \min\{a_i(z)^2, \lambda^2 s^2\} = s^2 \|\Pi_{\lambda B_\infty^d}(a(z)/s)\|^2.$$

This is nondecreasing in s . We have

$$\int_0^\infty U_z(s) \frac{ds}{s^2} = 2\lambda \sum_i |a_i(z)| \leq 2\lambda. \quad (31)$$

This is also the pointwise projection identity at 0 in direction $a(z)$ for the fixed cube λB_∞^d , with $\sigma = 1/s$. The same disjoint-interval calculation gives $\sum_n U_z(s_n)/s_n \leq 7\lambda$ along each branch. An uncovered point satisfies $2^{n/2} \text{diam}(B) \leq 512 U_z(s_n)/s_n$, while covered children have diameter at most $\text{diam}(B)/2$. The branch summation, starting at level 2 and using $\text{diam}(K_\varepsilon) \leq 4$, gives cost at most $C\lambda$. Finally, the $2d$ vertices are pairwise separated by at least $\sqrt{2}$, so $\gamma_2(K_\varepsilon) \geq c\sqrt{\log(ed)}$. ■

For the ℓ_1 ball, the construction of Wootters presented in [17, §4.4] and the construction in [18, §11.2.1] use global sparse nets, with the cost controlled by dyadic coordinate decompositions. Clipped coordinates and packing on their active supports already appear in [16, §8]; see also [14, §3.2]. The present example uses closely related sparse coordinate geometry within the general splitting construction of Section 3.

4.2 Minoration

Finally, we show that the same testing principle gives an elementary proof of the dual Sudakov inequality under (KL), shortening the information-theoretic proof in [1, Proposition 5.3]. For a symmetric convex body B , write $M(A, \varepsilon B)$ for the largest cardinality of a subset of A with pairwise distances greater than ε in the norm $\|\cdot\|_B$.

Proposition 11. *Let $T = -T$ be a convex body and denote its polar by $T^\circ = \{x : \sup_{t \in T} \langle x, t \rangle \leq 1\}$. Under (KL),*

$$w_\mu(T) \geq \frac{1}{64\sqrt{\kappa}} \sup_{\varepsilon > 0} \varepsilon \sqrt{\log M(B_2^d, \varepsilon T^\circ)}. \quad (32)$$

Proof. Choose $t_1, \dots, t_m \in B_2^d$, $m \geq 2$, separated by more than ε in $\|\cdot\|_{T^\circ}$, and put

$$\sigma = \frac{\varepsilon}{16w_\mu(T)}, \quad P_i = \mathcal{L}(t_i + \sigma X), \quad Q = \mathcal{L}(\sigma X).$$

We have $\mathbb{E}_{P_i} \|Y - t_i\|_{T^\circ} = \sigma \mathbb{E} \|X\|_{T^\circ} = \sigma w_\mu(T)$. The events $E_i = \{y : \|y - t_i\|_{T^\circ} \leq \varepsilon/2\}$ are pairwise disjoint by the separation of the t_i . The Markov inequality gives

$$P_i(E_i^c) \leq \frac{2\sigma w_\mu(T)}{\varepsilon} = \frac{1}{8}, \quad \mathcal{KL}(P_i \| Q) \leq \frac{\kappa}{\sigma^2}.$$

Applying Lemma 1, we obtain

$$\log m \leq \frac{16\kappa}{\sigma^2} = \frac{2^{12}\kappa w_\mu(T)^2}{\varepsilon^2}.$$

Taking square roots and then the supremum over $\varepsilon > 0$ proves the proposition. $\blacksquare$

Remark 7. Recent work also uses related information-theoretic tools. In [12, §§2.1, 2.4], operator norm Sudakov minoration for Gaussian chaos of order two is proved using adaptive Gaussian experiments and KL estimates against shifted reference laws.

References

- [1] R. Pathak and N. Zhivotovskiy. *Gaussian Width of Convex Sets via Integral Decompositions, Projections, and the Distribution of Intrinsic Volumes*. arXiv:2603.02714v2, 2026.
- [2] I. Zadik. *A Bayesian Proof and Interpretation of Talagrand’s Majorizing Measure Theorem*. arXiv:2605.30321v3, 2026.
- [3] R. Pathak and N. Zhivotovskiy. *A remark on the majorizing measures theorem for general processes*. arXiv:2606.03973v2, 2026.
- [4] M. Talagrand. *Upper and Lower Bounds for Stochastic Processes: Decomposition Theorems*. Second edition, Springer, 2021. Author version.
- [5] M. J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge University Press, 2019. doi:10.1017/9781108627771.
- [6] R. van Handel. *On the subgaussian comparison theorem*. arXiv:2512.18588v2, 2026.
- [7] J. Liu. *Simple and Sharp Generalization Bounds via Lifting*. arXiv:2508.18682v1, 2025.
- [8] R. M. Dudley. *The sizes of compact subsets of Hilbert space and continuity of Gaussian processes*. J. Functional Analysis 1 (1967), 290–330. doi:10.1016/0022-1236(67)90017-1.
- [9] X. Fernique. *Régularité des trajectoires des fonctions aléatoires gaussiennes*. In *École d’Été de Probabilités de Saint-Flour IV-1974*, Lecture Notes in Math. 480, Springer, 1975, 1–96. doi:10.1007/BFb0080190.
- [10] A. Song and M. Talagrand. *A very short proof of the subgaussian convex-order comparison theorem*. arXiv:2610.07876, 2026.
- [11] M. Talagrand. *Regularity of Gaussian processes*. Acta Math. 159 (1987), 99–149. doi:10.1007/BF02392556.
- [12] W. Bednorz, R. Martynek, and R. Meller. *Operator-norm Sudakov minoration for Gaussian chaos of order two*. arXiv:2609.23558v2, 2026.
- [13] F. Krahmer, S. Mendelson, and H. Rauhut. *Suprema of chaos processes and the restricted isometry property*. Comm. Pure Appl. Math. 67 (2014), 1877–1904. arXiv:1207.0235.
- [14] R. van Handel. *Chaining, interpolation, and convexity*. J. Eur. Math. Soc. 20 (2018), 2413–2435. doi:10.4171/JEMS/815.
- [15] R. van Handel. *Chaining, interpolation and convexity II: The contraction principle*. Ann. Probab. 46 (2018), 1764–1805. doi:10.1214/17-AOP1214.
- [16] M. Talagrand. *Majorizing measures: the generic chaining*. Ann. Probab. 24 (1996), 1049–1103. doi:10.1214/aop/1065725175.
- [17] J. Nelson. *Chaining introduction with some computer science applications*. Bull. Eur. Assoc. Theor. Comput. Sci. 120 (2016). Article.
- [18] S. Mendelson. *Geometric Aspects of Statistical Learning Theory — A Preliminary Draft*. ESI lecture notes, 2020. Lecture notes.
- [19] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Second edition, Cambridge University Press, 2026. Author version.
- [20] H. W. J. Zhang and C. Zhao. *Convex-Hull Instability of the γ_2 -Functional in L^p* . arXiv:2609.06614, 2026.