ONLINE KOMLÓS CONVERGES TO MEAN CURVATURE FLOW

NESTOR GUILLEN AND VLADIMIR A. KOBZAR

Dedicated to Robert V. Kohn, 1953-2026

Abstract. We determine the asymptotics of a game inspired by classic vector balancing problems in combinatorial discrepancy theory. In this game, which we call the online Komlós game, two players, Paul and Carol, update the state vector y in $\mathbb{R}^m$, initially placed at 0. At each round, Paul chooses freely a set of n vectors in the Euclidean unit ball, and Carol chooses, for each such vector, whether to leave it unchanged or reverse its sign. The resulting vectors are all added to y, and the game proceeds to a new round. After T rounds, the game ends, and the ℓ_∞ norm of the state vector y is determined. Paul's objective throughout the game is to maximize this norm, and Carol's objective is to minimize it. As T gets large, we establish that the leading order term of the value of this game is $\sqrt{T/2\tau}$, where τ is the extinction time of the unit cube in $\mathbb{R}^m$ under a curvature-based flow characterized by the values of m and n. When $n \geq m-1$, this flow is the mean curvature flow, and we show that $1/\sqrt{2\tau} = \Theta(\sqrt{\log m})$. Our results build upon the work of Kohn and Serfaty on deterministic games and mean curvature flow, combined with Banaszczyk's ℓ^2 analogue of the Beck-Fiala theorem. As the large T limit of the online Komlós game amounts to a localization of the classic Komlós problem, we hope this work can shed light on this and other vector balancing problems. Our results generalize to the version of the online Komlós game with the final value given by an arbitrary norm in $\mathbb{R}^m$.

1. Introduction

Discrepancy theory covers a wide array of problems and methods in combinatorics, geometry and analysis that involve approximation of a mathematical object by discrete elements [89, 111]. This field has been a long-standing focus of computer science and scientific computing, ranging from differential privacy and analog-to-digital conversion to causal inference and learning theory [8, 38, 40, 64, 67, 92]. One of the main problems in *combinatorial* discrepancy is vector balancing: given vectors $a_1, \ldots, a_k$ in a normed space, our objective is to choose signs $\varepsilon_1, \ldots, \varepsilon_k \in \{-1, 1\}$ in order to approximate the zero vector, i.e., minimize

$$\|\varepsilon_1 a_1 + \ldots + \varepsilon_k a_k\|.$$

We will focus on a version of this problem posed as a zero-sum game between two players, Paul and Carol.[1] The game is played for T rounds where Paul reveals a set of $k = nT$ vectors in $\mathbb{R}^m$ for $m \geq 2$ to Carol not all at once, but sequentially in T batches of size n.

Since the case when $T = 1$ corresponds to the classic open problem known as the Komlós conjecture, we will refer to our game as the *online Komlós game*, formally defined as follows. At round $t = 0$, we initialize the vector $y_t \in \mathbb{R}^m$ with $y_0 = 0$. In each subsequent round from $t = 1$ until the final round T, Paul chooses n vectors $a_{1,t}, \ldots, a_{n,t} \in \mathbb{R}^m$, each with at most unit Euclidean length, and reveals them to Carol. Then Carol irrevocably chooses n signs $\bar{\varepsilon}_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t}) \in \{-1, 1\}^n$, and we update

$$y_t = y_{t-1} + (\varepsilon_{1,t} a_{1,t} + \ldots + \varepsilon_{n,t} a_{n,t}).$$

We will write this update more compactly as $y_t = y_{t-1} + A_t \bar{\varepsilon}_t$ where A_t denotes the $m \times n$ matrix whose columns are given by the vectors $a_{1,t}, \ldots, a_{n,t}$. Paul's objective is to make the norm $\|y_T\|_\infty$ as large as possible, while Carol's objective is to make it as small as possible.

Paul is referred to as an *adaptive* adversary because he can choose each A_t based on Carol's choices of $\bar{\varepsilon}_\tau$'s at earlier times $\tau < t$. Accordingly, the value $K_T(m, n)$ of the online Komlós game is given by

$$K_T(m, n) := \max_{A_1} \min_{\bar{\varepsilon}_1} \ldots \max_{A_T} \min_{\bar{\varepsilon}_T} \|A_1 \bar{\varepsilon}_1 + \ldots + A_T \bar{\varepsilon}_T\|_\infty \tag{1.1}$$

Date: July 9, 2026.

[1] These names are mnemonics for *pusher* of vectors and *chooser* of signs, respectively, in the games introduced by Spencer.

where the feasible set of each A_t is given by all $m \times n$ matrices with columns of at most unit Euclidean length and each $\bar{\varepsilon}_t \in \{\pm 1\}^n$.

Kohn and Serfaty [78] established a connection between a version of this game and flow by mean curvature (see also [79]). Concretely, they found that their Paul and Carol game that is closest to ours (which they refer to as the time-dependent "inverse game"), is characterized in a continuum limit by the following nonlinear PDE

$$\partial_t u + \max\left\{\Delta u - \left(D^2 u \tfrac{\nabla u}{|\nabla u|}, \tfrac{\nabla u}{|\nabla u|}\right), 0\right\} = 0.$$

They also consider the continuum limit of several other games and arrive at various other PDE, such as

$$\partial_t u + \Delta u - \left(D^2 u \tfrac{\nabla u}{|\nabla u|}, \tfrac{\nabla u}{|\nabla u|}\right) = 0. \tag{1.2}$$

They obtain this last PDE using a game where Paul is restricted to choose batches forming orthonormal frames of $m-1$ vectors. The equation (1.2) is known as the level set formulation for mean curvature flow (MCF), since the level sets of the solution $u(\cdot, t)$ move with velocity equal to their mean curvature.

In this work we obtain a continuum limit for the online Komlós game, whose value is given by (1.1). Namely, we show an analogue of the result of Kohn and Serfaty holds in our setting, when the only condition on the vectors $a_{i,t}$ is that they all lie in the Euclidean unit ball. Accordingly, our first result shows that the leading order term of $K_T(m,n)$ as $T \to \infty$ is given in terms of the extinction time of the cube $[-1,1]^m$ under mean curvature flow. (In the interest of exposition we state here only special cases of the main theorems and defer the complete statements along with relevant definitions to Section 2.)

Theorem 1.1. *If $m \geq 2$ and $n \geq m-1$,*

$$\lim_{T\to\infty} \frac{K_T(m,n)}{\sqrt{T}} = \frac{1}{\sqrt{2\tau}},$$

where τ is the extinction time of $[-1,1]^m$ under mean curvature flow.

The complete result is recorded later in Theorem 2.1; it covers asymptotics for online vector balancing in any norm and batches of any size n, including $n = 1$. This result opens the door to studying the asymptotics of the value function $K_T(m,n)$ using tools from PDEs and geometry. Our second result characterizes how the long-time asymptotics in Theorem 1.1 grow with m.

Theorem 1.2. *The following holds as $m, n \to \infty$ with $n \geq m-1$*

$$(1 - o(1))\sqrt{\log m} \leq \lim_{T\to\infty} \frac{K_T(m,n)}{\sqrt{T}} \leq (\sqrt{2} + o(1))\sqrt{\log m}.$$

Equivalently, for τ the extinction time of $[-1,1]^m$ under mean curvature flow, we have

$$\frac{1-o(1)}{2\log m} \leq 2\tau \leq \frac{1+o(1)}{\log m}.$$

These estimates are based on a more general result covering centrally symmetric convex bodies, see Theorem 2.2 and Remark 2.3; they are obtained by means of comparison principle arguments, relating the solution of (1.2) to solutions of a linear heat equation. We believe our bounds are of interest independent of the online Komlós game. As we discuss in Section 1.3, smooth potentials, like the hyperbolic cosine and exponential weights, have been used extensively to prove *upper* bounds in offline and online vector balancing problems. On the other hand, our use of a potential (the heat equation solution) to establish a *lower* bound appears less common in this setting (one example is [23]). A similar approach has been used to establish lower bounds on regret in online learning, see for example [74–77].

1.1. **The classic and online Komlós problems.**

As noted earlier, when $T = 1$, our game reduces to a classic problem, which is subject to a long-standing conjecture attributed to János Komlós [105]:

The Komlós conjecture: $K_1(m,n) = O(1)$ as $m, n \to \infty$.

Without loss of generality, it is sufficient to resolve this conjecture when $n = m$ [28], [107, Lecture 5]. In the 90's, Banaszczyk established estimates for a broad class of vector balancing problems [10, 11], obtaining an $O(\sqrt{\log m})$ upper bound for $K_1(m,n)$ [11]. Paul can achieve the trivial lower bound of 1 on $K_1(m,n)$ by choosing an arbitrary orthonormal set of m vectors in $\mathbb{R}^m$, and Kunisky improved this lower bound to

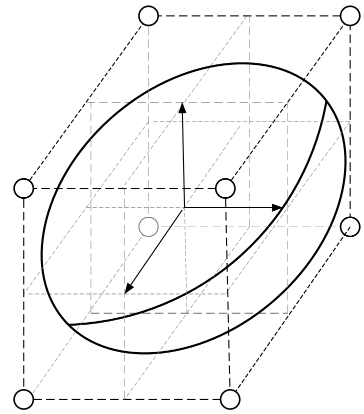

FIGURE 1. We have a convex body E and a frame with origin placed at the body's center. If A is the matrix whose columns represent the vectors in the frame, then as $\bar{\varepsilon}$ ranges over $\{-1,1\}^3$ we obtain 8 vectors $A\bar{\varepsilon}$ whose locations are marked above by white circles. In the classic Komlós problem Paul aims to choose the n (in this figure, $n=3$) column vectors of A so that the 2^n vertices $\{A\bar{\varepsilon}\}_{\bar{\varepsilon}}$ of the frame it defines all lie outside some λE, for a factor λ to be made as large as possible within the constraints for A.

$1+\sqrt{2}-o(1)$ as m gets large [82]. In recent work, Bansal and Jiang improved on the state of the art by proving an $\tilde{O}(\log^{1/4} n)$ upper bound in the Komlós setting [19].[2]

The Komlós conjecture is a statement about the asymptotics for $K_T(m,n)$ for $T=1$ as the dimension m and number of vectors n become large, while our Theorems 1.1 and 1.2 cover the large T regime, and consider the large m and n limit only after passing to the limit in T. It is an interesting possibility that the behavior of $u(x,t)$ solving (1.2) with $u(x,T)=\|x\|_\infty$ as $t\to T$ might encode information about the offline Komlós problem. We have tried without success to find such a connection. There might be a connection, or it might be that information about $K_1(m,n)$ is lost in the continuum limit.

Nonetheless, as random walk and potential-based iterative methods are used broadly in offline discrepancy problems, we hope our results will provide a helpful point of view for these methods. We also hope the PDE perspective proves helpful to other types of vector balancing problems, such as online problems with oblivious adversary or random vectors sampled from a fixed distribution. These problems are reviewed in Section 1.3.

Determining the dimensional asymptotics of the "large T asymptotics" is a localization of the original Komlós conjecture, as we now explain. It will be convenient to consider the Komlós problem for a general norm $\|\cdot\|_E$

$$\mathcal{K}_{n,1}(E) := \max_{A}\min_{\bar{\varepsilon}} \|A\bar{\varepsilon}\|_E,$$

where E stands for the unit ball for the norm (later, the case $\|\cdot\|_\infty$ will be approximated by other norms). From Paul's perspective, the classic Komlós problem consists (see Figure 1 above) in finding $a_1,\ldots,a_n$ so that no matter what Carol chooses for $\varepsilon_1,\ldots,\varepsilon_n$ Paul can reach, when starting from the origin, and in one step, as large as possible a multiple of ∂E. As such, Paul must consider all of E in choosing $a_1,\ldots,a_n$. Heuristically speaking, this is a global problem in E.

Contrast this with what happens in the online Komlós game with respect to $\|\cdot\|_E$ after a large number of steps (see Figure 2), that is, when we are considering for a large T,

$$\mathcal{K}_{n,T}(E) := \max_{A_1}\min_{\bar{\varepsilon}_1}\ldots\max_{A_T}\min_{\bar{\varepsilon}_T} \|A_1\bar{\varepsilon}_1+\ldots+A_T\bar{\varepsilon}_T\|_E.$$

Consider a vector y_{T-1} with the property that

$$\mathcal{K}_{n,T}(E) = \max_{A}\min_{\bar{\varepsilon}} \|y_{T-1}+A\bar{\varepsilon}\|_E,$$

in particular, we expect $\lambda := \|y_{T-1}\|_E$ to be large when T is large (see (3.1) later on). In this case, Paul finds y_{T-1} placed on a large rescaling $\lambda\partial E$ of ∂E, and he is trying to reach a set $(\lambda+h)\partial E$ for as large a

[2]The prefactor inside $\tilde{O}(\cdot)$ is polynomial in $\log\log n$, specifically $(\log\log n)^c$ for some fixed exponent $c>0$ uniformly in m.

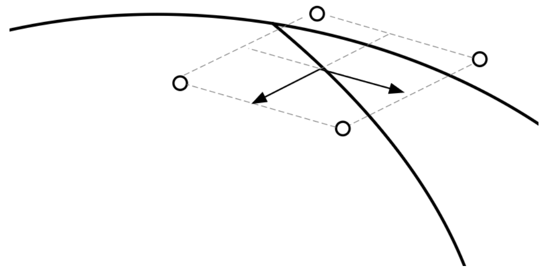

Figure 2. For the online problem ($m = 3, n = 2$ in this figure) we expect after a large number of steps t that y_t will lie on $\lambda\partial E$ for a large λ. Moreover, since the column vectors of A have length at most 1, the game is effectively zooming in to a small portion of ∂E. If the body E is smooth, the area where Paul and Carol play is well approximated by paraboloid tangent to ∂E at $\hat{y}_t := \|y_t\|_E^{-1} y_t$. Therefore, Paul's choice for the two vectors (and corresponding frame of 2^n points) is largely determined by the curvature of ∂E at $\hat{y}_t$.

value of $h > 0$ as possible. However, since the columns of A have ℓ_2 norm at most 1, the available positions for Paul which are all of the form $y_{T-1} + A\bar{\varepsilon}$ see only a relatively small region of $\lambda\partial E$ – since we expect λ to be large when T is large.

Then, if ∂E is smooth, the region of $\lambda\partial E$ immediately accessible to Paul from y_{T-1} (see Figure 2) will be approximately an ellipsoid (again, provided T and thus λ is large). This ellipsoid will be one passing through y_{T-1} and have the same curvature at y_{T-1} as ∂E, indicating that in this regime (T large) the curvature of ∂E governs how $\mathcal{K}_{n,T}(E)$ is changing with T. This heuristic picture is discussed further in Section 3.

1.2. **Notation.** For a vector x in $\mathbb{R}^n$, $\|x\|_p$ denotes its ℓ_p norm, and $|x| := \|x\|_2$ specifically denotes the Euclidean (ℓ_2) norm. For vectors x and y in $\mathbb{R}^m$, (x, y) denotes their Euclidean inner product. The vector e_i represents the i-th canonical basis vector of $\mathbb{R}^m$. Given a centrally symmetric body $E \subset \mathbb{R}^m$, $\|\cdot\|_E$ denotes the unique norm in $\mathbb{R}^m$ whose unit ball is equal to E. Its dual norm $\|\cdot\|_E^*$ is given by $\|z\|_E^* := \sup_{\|y\|_E \le 1}(z, y)$. For a matrix $M \in \mathbb{R}^{m\times m}$, its operator norm is denoted by $\|M\|$. Finally, the discrepancy of a matrix $A \in \mathbb{R}^{m\times n}$ is defined as

$$\operatorname{disc}(A) := \min_{\bar{\varepsilon}\in\{\pm 1\}^n} \|A\bar{\varepsilon}\|_\infty.$$

1.3. **Previous work on vector balancing.** The taxonomy of vector balancing problems is based on two primary characteristics. The first one specifies how the vectors are chosen. When Paul chooses vectors adversarially, as in our game, this setting is called the *worst-case* discrepancy. On the other hand, the *average case* discrepancy refers to the setting when the vectors are sampled independently from a distribution known to Carol. The second primary characteristic specifies when the vectors are revealed to Carol: in the *offline* setting Paul reveals all of them before Carol chooses any signs, while in the *online* setting, Paul reveals the vectors one-by-one and Carol irrevocably chooses the sign of each vector revealed in a given period, before the next vector is revealed [104]. Our online Komlós game interpolates between the purely online setting when Paul reveals one vector at a time and is the strongest adversary ($n = 1$), and the purely offline setting when Paul reveals all vectors at once and is the weakest adversary ($T = 1$).

Offline and online vector balancing problems have been studied in large part using randomized and/or potential-based iterative methods. Since it is plausible that some of these problems may be analyzed using PDE-based methods, we will now review some of the previous work on vector balancing, highlighting connections to those iterative methods. Due to space limits, we could not survey all potentially relevant recent discrepancy problems, such as Gaussian discrepancy [42], online geometric and interval discrepancy [23, 71], bad science matrices [2, 100, 109], or matrix discrepancy [12, 13, 83]. We refer the reader to the ICM survey by Bansal for a more comprehensive and in-depth treatment of discrepancy theory and related algorithms [16].

Offline worst case. One of the classic discrepancy minimization problems entails coloring the elements of a set U of size n with binary colors $\{\pm 1\}^n$ so that each member of a given collection of m subsets of U is colored in the most balanced way possible. This collection can be represented by the incidence matrix $A \in \{0,1\}^{m\times n}$, and the balance of coloring is measured by disc(A). For a closely related class of matrices: $A \in \mathbb{R}^{m\times n}$ and all $|A_{ij}| \le 1$ (we will refer to this setting as the *set coloring*), Spencer established that

$$\text{disc (A)} = O\big(\sqrt{n\log(2m/n)}\big)$$

by a partial coloring method using the binary entropy potential [105]. This result was also obtained by Gluskin using a different method also based on partial coloring [60], and Giannopoulos provided a simplified version of Gluskin's proof [55]. When $n \ge m$, this upper bound can be taken to be $5.32\sqrt{m}$, which led to its memorable "six standard deviations suffice" characterization. (Pesenti and Vladu [96] improved the constant prefactor in the set discrepancy setting from 5.32 to 3.68.) A matching $\Omega(\sqrt{m})$ lower bound in this setting is achieved by Hadamard matrices [3, 38].

Spencer also showed that if Carol can delete a constant fraction of vectors, i.e., set their colors to zero, then the Komlós conjecture holds for this modified problem [105]. Bandeira, Maillard and Zhivotovskiy [14] provided an alternative proof of this result based on the partial coloring theorem of Kashin [73], and yet another earlier proof of this result was given by Lonke [87]. On the other hand, Chasapis and Skarmogiannis showed that if Carol is prevented from using all possible colorings, specifically if the feasible set of $\bar{\varepsilon}$ is $S \subset \{\pm 1\}^m$ has at most $2^{\delta n}$ colorings for $\delta \in (1/m, 1)$, then the discrepancy in this modified problem is lower bounded as $\Omega\big(\sqrt{\log(e/\delta)}\big)$ [37], and Hajela established a similar result earlier [66]. Recently, Smirnov and Vershynin gave a Fisher information perspective on partial coloring based on the first Dirichlet eigenvalue of the hypercube [101].

Bansal developed the first polynomial-time algorithm for *partial* coloring that achieves Spencer's upper bound using a random walk with covariance determined by semidefinite programming (SDP) [15]. Subsequently, Lovett and Meka developed a simpler algorithm that achieved the same upper bound using a random walk inside the unit hypercube in a polytope $\mathbb{R}^m$ with certain discrepancy constraints [88]. Rothvoss, as well as Eldan and Singh, developed partial coloring algorithms that generalize to symmetric convex bodies beyond polytopes [48, 98]. Charikar, Newman, and Nikolov established that it is NP-hard to distinguish between set systems with discrepancy zero and $\Omega(\sqrt{m})$; thus we do not expect to be able to efficiently find an optimal coloring for set systems with an $o(\sqrt{m})$ discrepancy [35]. On the other hand, a closely related quantity called *hereditary discrepancy* of a matrix A, or heredisc(A), that upper bounds disc(A), can be efficiently approximated [90].

Bansal, Dadush, and Garg developed the first efficient algorithm based on a random walk with covariance determined by an SDP solution that achieved Banaszczyk's upper bound in the Komlós setting [17]. Levy, Ramadas and Rothvoss developed a deterministic algorithm based on the multiplicative weights potential that achieved this upper bound; this algorithm also achieves the $O(\sqrt{m})$ bound in the set discrepancy setting [85]. Subsequently, Bansal, Dadush, Garg and Lovett developed an algorithm that achieves Banaszczyk's bound for vector balancing in general convex bodies; their algorithm is based on a random walk that evolves in directions obtained by the Gram-Schmidt process [18]. Harshaw, Sävje, Spielman, and Zhang improved the variance bounds of the signed vectors sums obtained in this process [67], and Bednorz and Godlewski used this Gram-Schmidt walk to improve (i) the constant prefactor in Banaszczyk's $O(\sqrt{\log m})$ upper bound in the Komlós setting and (ii) the discrepancy bounds in the smoothed analysis of the Komlós conjecture with Gaussian noise (this setting is further discussed below) [29]. Lastly, Bansal, Laddha and Vempala gave a deterministic algorithm based on a time-dependent barrier function that achieves all the bounds described above – Spencer's set coloring bound $O\big(\sqrt{n\log(2m/n)}\big)$, Banaszczyk's upper bounds in the Komlós setting – as well as certain bounds in the spectral hypergraph and Gaussian random matrix settings [24].

Bansal, Jiang, Meka, Singla and Sinha applied the smoothed analysis of Spielman and Teng [108] to show that the Komlós conjecture holds asymptotically almost surely (a.a.s.) when $n = \omega(m\log m)$ and A is perturbed by adding small Gaussian noise [22] (see also [21] for the smoothed analysis of the prefix discrepancy). Aigner-Horev, Hefetz and Trushkin established a similar result when A is perturbed by Rademacher noise [1].

Offline average case. While the Komlós conjecture remained unresolved, a substantial amount of work was dedicated to estimating discrepancy of special classes of random and pseudorandom matrices. For Gaussian matrices $A \in \mathbb{R}^{m\times n}$ with each $A_{ij} \sim N(0,1)$ independently, Costello established that disc$(A) =$

$\Theta(2^{-n/m}\sqrt{n})$ a.a.s. when m is fixed [44], and Turner, Meka and Rigollet established the same result when $\omega(1) = m = o(n)$ [112].[3] In the Gaussian matrix setting, Chandrasekaran and Vempala upper bounded discrepancy as $O(\sqrt{n\log(m/n)})$, modulo a polylog term in m and n, when $2n \leq m \leq 2^n$ [34]. Pesenti and Vladu gave an algorithm based on the Stieltjes regularizer of the max function that was used to prove the Komlós conjecture for certain pseudorandom matrices, as well as for random orthogonal and random Gaussian matrices [96].

In the discrete random matrix setting, Ezra and Lovett showed that when the columns of $A \in \mathbb{R}^{m\times n}$ are sampled from s-sparse vectors in $\{0,1\}^m$ uniformly and $m \geq n \geq s$, $\mathrm{heredisc}(A) = O(\sqrt{s\log s})$ and when $n \gg m^s$, $\mathrm{heredisc}(A) = O(1)$ with high probability [50]. Hoberg and Rothvoss showed $\mathrm{disc}(A) \leq 1$ with high probability when $A_{ij} \sim \mathrm{Bernoulli}(p)$ independently provided that $n = \Omega(m^2\log m)$ and $mp = \Omega(\log n)$ [68]. Franks and Saks showed that $\mathrm{disc}(A) \leq 2$ with high probability for a rather broad class of random s-sparse matrices when $n = \Omega(m^3\log^2 m)$ [53]. Altschuler and Niles-Weed showed $\mathrm{disc}(A) = O(1)$ with high probability when $A_{ij} \sim \mathrm{Bernoulli}(p)$ independently and $n \geq Cm\log m$ where p may be varying with n; they also provided a discrepancy upper bound for a broad class of Poisson random matrices [4].

Online worst case with adaptive adversary. As noted earlier, in the context of online discrepancy, Paul is called an adaptive adversary whenever he is able to choose the vectors based on Carol's choices of signs at earlier times. Spencer showed that in the online version of the set discrepancy problem, where Paul chooses a_t with $\|a_t\|_\infty \leq 1$ adaptively, the discrepancy is $\Theta(\sqrt{T\log T})$; the upper bound is established using the cosh potential, while the lower bound is established by analyzing the distribution of certain hitting times of a one-dimensional random walk [107, Lecture 4] (see also [106]). Haghtalab, Roughgarden and Shetty gave an algorithm that guarantees an $\tilde{O}(\log^2(mT/\sigma))$ upper bound on discrepancy in a modified version of our online Komlós game with $n = 1$ where the adversary selects the vectors from an adaptive σ-smooth isotropic distribution where density is upper bounded pointwise by $1/\sigma$ times the density of the uniform distribution. Their algorithm uses the cosh-based potential introduced by Bansal, Jiang, Meka, Singla and Sinha in the online average case discussed below [20, 65]. Without smoothing our online Komlós game with $n = 1$ appears trivial: by equation (3.1),

$$1/\sqrt{2} \leq K_t(m,1)/\sqrt{t} \leq 1,$$

uniformly in m (see also the discussion of Game 1 in [104]). Perhaps due to this, much of the recent online vector discrepancy literature focused on the average case discrepancy, as well as adversaries that are not adaptive, the so-called *oblivious* adversaries.

Online average case. When the vectors are sampled from the uniform distribution over $\{\pm 1\}^m$ independently and revealed to Carol online, Bansal and Spencer gave a strategy based on an inverse polynomial potential that guarantees with high probability an $O(\sqrt{m})$ discrepancy, as well as another strategy based on the cosh potential that guarantees an $O(\sqrt{m}\log m)$ prefix discrepancy [25] (the *prefix discrepancy* controls the ℓ_∞ norm of the partial sums y_t uniformly across time, not just at the final time). When the vectors are sampled independently from an arbitrary distribution over $[-1,1]^m$ with potentially non-independent entries, Bansal, Jiang, Singla and Sinha gave an $O(m^2(\log T + \log m))$ online prefix discrepancy bound using the cosh potential [23]. Improving on this result, Bansal, Jiang, Meka, Singla and Sinha gave an $O(\log^4(mT))$ bound on online prefix discrepancy when vectors are sampled independently from an arbitrary distribution on the unit ℓ_2 ball in $\mathbb{R}^m$ [20].

A special case of the Komlós conjecture is the *Beck-Fiala conjecture*: for any $A \in \{0,1\}^{m\times n}$ with column sparsity of at most s, $\mathrm{disc}(A) = O(\sqrt{s})$ [28]. In this setting, Altschuler and Tikhomirov established that there exists an asymptotic gap between the discrepancy in the online and offline vector balancing of independent uniformly random s-sparse vectors [5]. Finally, when the vectors are sampled independently from the standard m-dimensional Gaussian and revealed to Carol online, Fiedler, Jackson, Lacker and Niles-Weed established a $\Theta(\sqrt{m})$ discrepancy estimate in expectation; their lower bound applies more generally to all random vectors with i.i.d. mean zero and variance 1 entries with finite fourth moments [51]. Although their methods are grounded in mean-field stochastic control, their use of a discrete value function defined by dynamic programming is similar to ours.

Online worst case with oblivious adversary. In the oblivious setting, Paul still reveals the vectors to Carol sequentially. However, he cannot use Carol's earlier choices of signs when he selects a vector in a given

[3] Thus, $\mathrm{disc}(A) = O(1)$ a.a.s. as long as $n \geq c\,m\log m$ for a some uniform constant c. By standard Gaussian properties, each column $A_{:i}$ has unit ℓ_2 norm a.a.s., and by the union bound $A/\sqrt{m}$ satisfies the hypothesis of the Komlós conjecture a.a.s.

round. Since every deterministic algorithm for choosing signs has the same worst case discrepancy against an oblivious adversary as against the corresponding adaptive adversary (whenever the oblivious adversary correctly guesses the worst case pattern of vectors in the adaptive case) Carol is permitted to use randomized policies, and the discrepancy is evaluated with high probability with respect to Carol's randomization.

In the online Komlós setting with an oblivious adversary, Kulkarni, Reis and Rothvoss achieved the state-of-the-art $O\big(\sqrt{\log(\min(m,T))}\big)$ discrepancy upper bound and $\Theta\big(\sqrt{\log T}\big)$ prefix discrepancy upper and lower bounds, modulo a term that depends on the probability of attaining this bound [81]. Their upper bound strategy depends on the entire series vectors $a_1, \dots a_{t-1}$ (and therefore does not appear to be Markovian). Earlier, Alweiss, Liu and Sawhney achieved an $O(\log mT)$ prefix discrepancy bound with high probability using a "self-balancing random walk" algorithm that is Markovian - it depends on the current time t and the inner product of the partial sum $\sum_{\tau=1}^{t-1} \varepsilon_\tau a_\tau$ with the vector a_t revealed to Carol in the current period [6].

1.4. **Mean curvature flow and the level set PDE.** The nonlinear PDE (1.2) is known as the level set formulation of mean curvature flow, and it is one of several alternative descriptions of this geometric evolution. The mean curvature flow appears in many areas of pure and applied mathematics, receiving thorough and complementary treatments from the perspectives of geometric measure theory [31, 110], the theory of nonlinear parabolic equations [41, 49, 56], and geometric analysis [43], among others.

Brakke's mean curvature flow. The mean curvature flow was introduced by Brakke in his doctoral dissertation [31], and accordingly sometimes it is known as Brakke's mean curvature flow. Informally speaking, a time-evolving family of closed hypersurfaces $S_t \subset \mathbb{R}^m$ (enclosing domains E_t, so $S_t = \partial E_t$) is said to be evolving by mean curvature flow if the inner normal velocity of S_t at any point $p \in S_t$ is equal to the sum of the principal curvatures of S_t at p.

The informal description above masks a number of deep analytical and geometrical difficulties revolving around the right definition of such a flow, and the nature of the singularities that may form over time even if one starts the flow from a smooth hypersurface. Brakke's original work [31] falls within the framework of geometric measure theory, working with varifolds to formulate a notion of solution for this flow. For a modern treatment of the mean curvature flow from this perspective, see Tonegawa's monograph [110].

Huisken's theorem and extinction times. A celebrated Theorem of Huisken [69] says that if the initial surface S_0 is the boundary of a convex body E_0, then the surface remains convex for the duration of the mean curvature flow. To be exact, Huisken showed there is a positive time $\tau = \tau(E_0)$ and a mean curvature flow S_t defined for $t \in [0, \tau)$ and starting from S_0, and such that S_t is the boundary of a smooth convex set for every $t \in (0, \tau)$, and that $S_t \to \{x_*\}$ in the Hausdorff topology to some point $x_* \in \mathbb{R}^m$.

In Huisken's work the flow is represented via a map $F : \mathbb{S}^{m-1} \times [0, \tau) \to \mathbb{R}^m$ such that $F(\cdot, t)$ gives a parametrization of S_t, the evolution of the map $F(\cdot, t)$ with t is governed by a nonlinear parabolic equation for the map F. As $t \to \tau$, the map $F(\cdot, \tau)$ converges to a constant (i.e. to the point to which the surfaces are shrinking to as $t \to \tau$). The time τ is known as the extinction time of S_0, or equivalently, of E_0 (see also (2.1)). It defines a functional over the set of convex bodies. In general, there is no explicit expression for the extinction time $\tau(E)$ for a given convex body E (a notable exception is a ball). We note that work of Giga and Yama-uchi [59] provides an estimate for the extinction time in terms of the Cheeger constant of E, which is a fundamental quantity in geometry related to the spectral gap of the Laplace-Beltrami operator [39].[4]

The level set formulation. Level set formulations are a way of describing hypersurfaces S of $\mathbb{R}^m$ in terms of a scalar function u, "the level set function", such that $S = \{u = 0\}$. In this way, one can describe geometric quantities of S in terms of an expression involving u. For instance, if $x \in S$, then the normal vectors to S are given by $\pm \nabla u / |\nabla u|$. Note also that a family of hypersurfaces S_t varying in time can be represented by a time-dependent level set function $u(x, t)$. For a general introduction to the level set method, see the book by Giga [57] and the book by Fedkiw and Osher [94].

A standard computation shows that if $u(x, t)$ is a smooth level set function and $\nabla u \neq 0$, then for each λ the level set $S_t(\lambda) = \{x : u(x, t) = \lambda\} = \partial\{x : u(x, t) < \lambda\}$ is evolving in time with an inner normal velocity

[4]For purposes of the Giga and Yama-uchi estimate, we include the portion of the perimeter on the hypercube boundary in the definition of the Cheeger constant,(i.e., the ambient set is $\mathbb{R}^m$). Therefore, the estimates of the Cheeger constant when the hypercube boundary is excluded from the perimeter are not applicable in the Giga and Yama-uchi setting (i.e., when the ambient set is the hypercube itself; cf. Section 5.5 in [91]).

V at $x \in S_t(\lambda)$ given by

$$V = \frac{\partial_t u}{|\nabla u|}. \tag{1.3}$$

This brings us to level set equations which are also parabolic[5] PDEs. These are equations of the form

$$\partial_t u + |\nabla u| F(D^2 u, \nabla u) = 0,$$

where $F(D^2u, \nabla u)$ is a degenerate parabolic operator, meaning that for a fixed p we have $F(M,p) \leq F(M',p)$ whenever $M' - M$ is a positive semi-definite matrix. In light of the formula (1.3), the significance of the parabolic PDE for u is that its level sets are moving with normal velocity equal to $F(D^2u(x), \nabla u(x))$. The level set equation is in addition called geometric if $F(M,p) = F(M + sp \otimes p, p)$ for every M, $p \neq 0$, and $s \in \mathbb{R}$. This means that the value of $F(D^2u(x), \nabla u(x))$ at some $x = x_0$ is determined by the shape of the level set of $\{u = u(x_0)\}$.

As an important example, the operator F giving the level set formulation of mean curvature flow is

$$F(D^2u, \nabla u) = \mathrm{tr}(\Pi(\nabla u)(D^2 u)\Pi(\nabla u))/|\nabla u| = \mathrm{div}\left(\tfrac{\nabla u}{|\nabla u|}\right),$$

where for $p \neq 0$, $\Pi(p)$ denotes the $m \times m$ matrix representing the orthogonal projection onto the hyperplane normal to p. This matrix is given by $(\Pi(p))_{ij} = \delta_{ij} - |p|^{-2} p_i p_j$. Then, the level set PDE for mean curvature flow is

$$\partial_t u + |\nabla u|\, \mathrm{div}\left(\tfrac{\nabla u}{|\nabla u|}\right) = 0.$$

These equations are fully nonlinear parabolic equations, and so they fall within the scope of the theory of viscosity solutions. The level set formulation was first studied by Evans and Spruck [49] and Chen, Giga, and Goto [41]. We also mention work of Giga, Goto, Ishii, and Sato [56] dealing with general flows and which will also be essential here. As discussed earlier, Kohn and Serfaty [78] found the level set PDE for mean curvature flow arises as a continuum limit in vector balancing games. Kohn and Serfaty subsequently extended the deterministic games framework to general fully nonlinear parabolic and elliptic equations [80], and Imbert and Serfaty extended it to include integral curvature flows [70]. Concurrently, work of Peres, Schramm, Sheffield, and Wilson [95] touched on parallel themes as they established a relationship between tug-of-war games and the infinity Laplacian. There is a 1-parameter family of problems connecting the infinity Laplacian [95] and mean curvature flow [78]: they are two extremes in the class of evolution p-Laplacian equations ($p \in [1, \infty]$). We also mention works of Charro, García Azorero, and Rossi [36], as well of Gonzalvez, Miranda, Rossi, and Ruiz-Cases [61] involving two-player zero-sum probabilistic games.

More general flows. One may evolve $S_t = \partial E_t$ according to some other function of the principal curvatures of S_t, leading to parabolic level set PDEs as above, as well as to PDE's for maps $F : \mathbb{S}^{m-1} \to \mathbb{R}^m$ in the smooth setting. We mention in particular work of Andrews [7], which considered general geometric flows starting from convex initial data. An important result in [7] is the preservation of convexity under natural structural assumptions on the flow (i.e. structural assumptions on the function of principal curvatures used). This is important in defining extinction times for other flows besides mean curvature.

One possibility is the flow where the normal velocity equals the sum of the largest $n \wedge (m-1)$ principal curvatures of S_t. The resulting evolution also produces a family of convex bodies that shrink down to a point at some finite time $\tau_n(K)$. When $n \geq m-1$ this is the usual mean curvature flow, and when $n < m-1$ we refer to this as flow by the largest n principal curvatures. Flows with $n = 1$ have been shown to arise in stochastic control problems [84] (see next paragraph).

Stochastic optimization interpretation. Prior to Kohn and Serfaty's deterministic game setting, various works showed mean curvature flow captures the reachability set for certain stochastic control problems. We mention here the works of Soner and Touzi [102,103], of Buckdahn, Cardaliaguet, and Quincampoix [32]. More recently, Larsson and Ruf [84] found that flow by the minimum curvature describes the largest (deterministic) amount of time that a certain martingale can be kept inside a given convex body.

Viscosity solutions and fully nonlinear equations. Viscosity solutions are a powerful tool when dealing with value functions for games or control problems involving in continuum settings. As we have known since works of Bellman [30] and Pontryagin [97], a sufficiently smooth value function will solve a PDE

[5]For brevity, throughout this paper we use "parabolic" to refer to "backward parabolic equations", as we will only be working with backward parabolic equations, in accordance to the structure of the online Komlós game.

associated to the problem, the Hamilton-Jacobi-Bellman equation. However, value functions are typically not smooth everywhere, and one must deal with generalized solutions. Among several notions of "weak solutions" to such nonlinear equations, viscosity solutions stand out for producing a complete theory of existence and uniqueness, and for correctly characterizing the weak formulation of the PDE for non-smooth value functions. Just as important is the role of viscosity solutions in the context of continuous limits, where we have Barles and Souganidis [27] celebrated theorem on the convergence of finite difference schemes to the unique viscosity solution of a PDE.

Viscosity solutions are built around a natural monotonicity property of the nonlinear operator, and which is directly related to the maximum principle. This monotonicity, sometimes known as the Global Comparison Property (GCP), imposes severe restrictions on the class of operators. In particular, Courrège showed that a linear operator in $\mathbb{R}^n$ satisfying the GCP is essentially the infinitesimal generator of a Markov process [45]. The corresponding result for nonlinear operators was obtained by Guillen and Schwab [62, 63], where they show that a nonlinear operator having the GCP can be expressed as a min-max of linear operators of the type obtained by Courrège.

For monographs on viscosity solutions we recommend the book of Fleming and Soner [52] and the book of Bardi and Capuzzo-Dolcetta [26], both of which also cover the perspective from optimal control. We also recommend the lecture notes by Silvestre [99] as well as lecture notes by Calder [33]. Last but not least we mention the forthcoming monograph by Giga and Liu [58], which covers geometric evolution equations arising from deterministic discrete two-person games.

2. Statement of the main results

In this section, we state the main results of this paper in full generality, containing each of the special theorems discussed in the Introduction (Section 1). For the rest of the paper E will denote a centrally symmetric convex body in $\mathbb{R}^m$, $m \geq 2$. This body defines a norm, which we denote by $\|\cdot\|_E$. Now, for any $T, n \in \mathbb{N}$ we define a functional over such convex bodies by

$$\mathcal{K}_{n,T}(E) := \max_{A_1} \min_{\bar{\varepsilon}_1} \ldots \max_{A_T} \min_{\bar{\varepsilon}_T} \|A_1\bar{\varepsilon}_1 + \ldots + A_T\bar{\varepsilon}_T\|_E,$$

where for each $t = 1 \ldots, T$, the minimum is taken over all $\bar{\varepsilon}_s \in \{-1, 1\}^n$ and the maximum is taken over all $m \times n$ matrices A whose columns have at most unit Euclidean length each.

As discussed in the introduction, given E and $n \geq 1$, there is well defined flow E_t at least for small positive times, $E_0 = E$ and ∂E_t moves with normal velocity equal to the sum of the largest $n \wedge (m-1)$ principal curvatures of ∂E_t. The sets E_t are all centrally symmetric convex bodies and the flow is defined up until some maximum interval of existence. Then, we define the *extinction time* for E with respect to the flow by

$$\tau_n(E) := \sup \left\{ t \;\middle|\; \begin{array}{l} \text{There exists a flow starting from } E \text{ with normal velocity equal to the sum of} \\ \text{the largest } n \wedge (m-1) \text{ principal curvatures, and it is non-vanishing in } (0,t). \end{array} \right\}. \tag{2.1}$$

For such a time, it is known that $E_t \to \{0\}$ in the Hausdorff distance as $t \to \tau_n(E)$.

Theorem 2.1. *For any centrally symmetric convex body E, we have*

$$\lim_{T\to\infty} \frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}} = \frac{1}{\sqrt{2\tau_n(E)}}.$$

where τ_n is the extinction time defined in (2.1).

Theorem 2.2. *For any centrally symmetric convex body E, if $n \geq m-1$, we have*

$$\frac{1}{l^2(E)} \leq 2\tau_n(E) \leq \frac{2}{C^2(E)M_m^2}.$$

where

$$C(E) := 1/\max_i \|e_i\|_E^*,$$

M_m is the expectation of the largest entry, and $l(E)$ is the expected $\|\cdot\|_E$ norm, of the standard Gaussian random vector $N(0, I_m)$. Equivalently,

$$\frac{1}{\sqrt{2}} C(E) M_m \leq \lim_{T\to\infty} \frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}} \leq l(E).$$

Remark 2.3. *$l(E)$ is equal to the Gaussian width of the unit ball of the dual norm $\|\cdot\|_E^*$, and for the ℓ_∞ norm, the leading order term of this quantity is $\sqrt{2\log m}$ as m gets large. When E is the unit ball of any ℓ_p norm, $C(E)$ is 1, and the leading order term of M_m is $\sqrt{2\log m}$ [54, §2.3.2] (see also [47, 72, 93] for nonasymptotic bounds).*

The foregoing theorems are proved in Sections 6 and 7. The next section provides heuristics and overview of the proofs, while Sections 4 and 5 as well as the Appendices set forth definition and results used in the proof of Theorem 2.1 in Section 6.

3. Heuristics and overview of the proofs

The connection between Spencer's pusher-chooser game and mean curvature flow was first made clear in Kohn and Serfaty's work [78]. In their setting, Paul chooses an orthonormal frame of $m-1$ vectors. This already produces mean curvature flow in the limit. Here we show for the same continuum limit holds when the constraint is that the vectors lie in the Euclidean unit ball (and in particular one can take any $n \geq m$ vectors). Ultimately, this is made possible by an estimate of Banaszczyk for discrepancy with respect to the ℓ^2 norm [9].

Elementary strategies. To explain in rough terms the appearance of mean curvature, it helps to discuss some elementary strategies for Paul and Carol, which guarantee the following bounds at all T:

$$\sqrt{c(m,n)T} \leq K_T(m,n) \leq \sqrt{\min(n,m)\,T} \tag{3.1}$$

where $c(m,n) := \min(n,m-1)/(\min(n,m-1)+1)$.

A possible strategy for Paul is to choose the columns of $a_{t,1},\ldots,a_{t,n^*}$ to be $n^* := \min(n,m-1)$ orthonormal vectors that are orthogonal to y_{t-1} in a n^*+1 coordinate subspace of $\mathbb{R}^m$ that stays fixed during the game, e.g., $\operatorname{span}\{e_1,\ldots,e_{n^*+1}\}$, and set any remaining vectors to zero. This choice means that regardless of Carol's choice of signs,

$$|\varepsilon_{t,1}a_{t,1}+\ldots+\varepsilon_{t,n}a_{t,n}|^2 = |\varepsilon_{t,1}a_{t,1}|^2+\ldots+|\varepsilon_{t,n}a_{t,n}|^2 = n^*.$$

(As a reminder, in our notation, $|\cdot|$ refers to the ℓ_2 norm.) Therefore, Paul can make sure that $|y_t|^2 = |y_{t-1}|^2 + n^*$ at every step t, and thus

$$|y_t|^2 = t\,n^*,\ \forall\, t.$$

Since $|y_t| \leq \sqrt{n^*+1}\,\|y_t\|_\infty$, one concludes that

$$\|y_t\|_\infty \geq \sqrt{c(m,n)t},\ \forall\, t,\ c(m,n) = n^*/(n^*+1).$$

Thanks to Banaszczyk's theorem (see Theorem 5.1), whatever Paul chooses for the vectors $a_{t,1},\ldots,a_{t,n}$, there exists a choice of signs ε_t for Carol such that

$$|\varepsilon_{t,1}a_{t,1}+\ldots\varepsilon_{t,n}a_{t,n}|^2 \leq m.$$

On the other hand, when the signs $\varepsilon_{i,t}$ are independently sampled uniformly at random, the expected value $\mathbb{E}|y_t|^2 \leq nt$, and by the monotonicity of the ℓ_p norm, $\mathbb{E}\|y_t\|_\infty \leq \sqrt{nt}$. Therefore, by a probabilistic argument, there exists a deterministic policy of choosing ε_t so that

$$|y_t|^2 \leq |y_{t-1}|^2 + \min(m,n), \Rightarrow \|y_t\|_\infty \leq \sqrt{t\min(m,n)}.$$

These strategies guarantee the bounds in (3.1).

Appearance of the mean curvature. There is a sort of generalization of the above argument for a non Euclidean norm, and it serves to illustrate the emergence of the mean curvature in the vector balancing context. Instead of the norm $\|\cdot\|_\infty$ let us consider a general norm $\|\cdot\|$ whose unit ball E is assumed to be smooth.

Observe that $y \mapsto \|y\|^2$ is a 2-homogeneous function, which is smooth away from 0, and in particular, the second derivatives of $\|\cdot\|$ are 0-homogeneous and their third derivatives are negatively homogeneous. This means that if y_{t-1} is very large we expect $\|\cdot\|^2$ to be close to its second Taylor expansion in say, a ball of radius 1 centered at y_{t-1}.

Then let us assume $\|y_{t-1}\|$ is large (as will be the case when t is large), and consider the unit vector $\hat{y}_{t-1} := y_{t-1}/\|y_{t-1}\|$ as well as the outer normal vector $n(\hat{y}_{t-1})$ at $\hat{y}_{t-1}$, which lies on ∂E with unit dual norm $\|n(\hat{y}_{t-1})\|^* = 1$. Using Euler's identity for 1-homogeneous functions, one can check that

$$\nabla \|\cdot\|^2(y_{t-1}) = 2\|y_{t-1}\| n(\hat{y}_{t-1}),\ D^2\|\cdot\|^2(y_{t-1}) = 2\mathrm{II}_{\partial E}(\hat{y}_{t-1}) + 2n(\hat{y}_{t-1}) \otimes n(\hat{y}_{t-1}),$$

where $\mathrm{II}_{\partial E}(y)$ denotes the second fundamental form of ∂E at a point $y \in \partial E$. Then, we can use the second order Taylor expansion for $\|\cdot\|^2$ near y_{t-1} to estimate $\|y_{t-1} + A\bar{\varepsilon}\|^2$,

$$\begin{aligned}\|y_t\|^2 &= \|y_{t-1} + A\bar{\varepsilon}\|^2 \\ &= \|y_{t-1}\|^2 + 2\|y_{t-1}\|(n(\hat{y}_{t-1}), A\bar{\varepsilon}) + (\mathrm{II}_{\partial E}(\hat{y}_{t-1})A\bar{\varepsilon}, A\bar{\varepsilon}) + (A\bar{\varepsilon}, n(\hat{y}_{t-1}))^2 + \mathrm{O}(\|y_{t-1}\|^{-1}).\end{aligned}$$

Paul can always make sure the gradient term vanishes regardless of Carol's choice. Paul simply needs to choose all the columns of A to be orthogonal to $n(\hat{y}_{t-1})$. In this case we do not see the components of $D^2\|\cdot\|^2(y_{t-1})$ in the directions parallel to $n(\hat{y}_{t-1})$, and we get

$$\|y_t\|^2 = \|y_{t-1}\|^2 + (\mathrm{II}_{\partial E}(\hat{y}_{t-1})A\bar{\varepsilon}, A\bar{\varepsilon}) + \mathrm{O}(\|y_{t-1}\|^{-1}).$$

Now, thanks to Banaszczyk's theorem the second term in the right hand side can be understood and even written in simple terms (this is explained later in Remark 5.2). In particular, if $n \geq m-1$ it can be shown that

$$\max_{A_t} \min_{\varepsilon_t} \|y_{t-1} + A_t\bar{\varepsilon}_t\|^2 = \|y_{t-1}\|^2 + H(\hat{y}_{t-1}) + \mathrm{O}(\|y_{t-1}\|^{-1}),$$

where $H(y)$ denotes the mean curvature of ∂E at $y \in \partial E$. This computation suggests that for large values of t the evolution of $\|y_t\|^2$ is dominated by the values of the mean curvature of ∂E, and that, accordingly, the large t asymptotics for $\|y_t\|$ might be governed by the mean curvature of ∂E.

3.1. **Overview of the proofs.** We will now give an overview of the proofs of Theorem 2.1 (which contains Theorem 1.1 as a special case) and of Theorem 2.2 (which contains Theorem 1.2 as a special case).

Overview of Theorem 2.1 on large T asymptotics for $\mathcal{K}_{n,T}(E)$. This is the more general counterpart to Theorem 1.1. The proof centers on a value function $v(y,s)$ that encodes $\mathcal{K}_{n,T}(E)$ for a general centrally symmetric convex body E. To facilitate the PDE analysis, we denote the steps s by negative integers with the starting time $-T$ and the final time 0. At the start, it will be convenient to consider the case where E has a smooth (C^2) boundary, with the general case handled via a stability and approximation argument at the very end of the proof. As in Kohn and Serfaty [78], we will show that this value function, properly rescaled, solves what amounts to a finite difference scheme for a nonlinear parabolic PDE.

Definition 3.1. *For a centrally symmetric convex body E in $\mathbb{R}^m$ with $m \geq 2$ and $n \geq 1$, we define the value function $v : \mathbb{R}^m \times \mathbb{Z}_- \to \mathbb{R}_+$ by the relation*

$$v(y,s) := \max_{A_s} \min_{\bar{\varepsilon}_s} \ldots \max_{A_{-1}} \min_{\bar{\varepsilon}_{-1}} \|y + A_s\bar{\varepsilon}_s + \ldots + A_{-1}\bar{\varepsilon}_{-1}\|_E,\ y \in \mathbb{R}^m,\ s = 0, -1, \ldots,$$

where for each $\tau = s, \ldots, -1$, the minimum is taken over all $\bar{\varepsilon}_\tau \in \{-1,1\}^n$ and the maximum is taken over all $m \times n$ matrices A whose columns have at most unit Euclidean length each.

Observe that $\mathcal{K}_{n,T}(E) = v(0,-T)$. Accordingly, what we are after is the limit $v(0,-T)/\sqrt{T}$ as $T \to \infty$. We will compute this limit by showing there is a continuum limit for a family of rescalings of v. For $\delta > 0$, we define $w_\delta : \mathbb{R}^m \times (\delta^2\mathbb{Z}_-) \to \mathbb{R}$ by

$$w_\delta(x,t) = \delta v(x/\delta, t/\delta^2).$$

Choosing $\delta = 1/\sqrt{T}$, we have

$$\lim_{T\to\infty} \frac{v(0,-T)}{\sqrt{T}} = \lim_{T\to\infty} w_{\frac{1}{\sqrt{T}}}(0,-1).$$

We will show that w_δ is a solution to the (nonlinear) finite difference scheme

$$\tfrac{1}{\delta^2}(w_\delta(x,t) - w_\delta(x,t-\delta^2)) + \max_A \min_{\bar{\varepsilon}} \left\{\delta^{-2}(w_\delta(x+\delta A\bar{\varepsilon},t) - w_\delta(x,t))\right\} = 0. \tag{3.2}$$

Thus, determining the large T asymptotics for $\mathcal{K}_{n,T}(E)$ is reduced to determining whether the above finite difference scheme approximates a PDE in the $\delta \to 0$ limit. In other words, we must determine whether w_δ

converges to the solution of a PDE as $\delta \to 0$. This entails, on one hand, obtaining estimates for $w_\delta(x,t)$ that are uniform in δ, and, on the other hand, identifying the limiting PDE itself. The estimates are obtained first for v (Section 4) by taking advantage of the smoothness of E and the dynamic programming principle, and then transferred to w_δ by rescaling.

We identify the limiting PDE in Section 5. First, we show that as $\delta \to 0$ the finite difference scheme captures only derivatives of order at most 2, that is, for any fixed C^2 function ϕ we have

$$\max_A \min_{\bar\varepsilon} \left\{\delta^{-2}(\phi(x+\delta A\bar\varepsilon) - \phi(x))\right\} = \tfrac{1}{2} H_{\delta,n}(D^2\phi(x), D\phi(x)) + o(1),$$

as $\delta \to 0$. Here, $H_{\delta,n}$ is given by

$$H_{\delta,n}(M,p) = \max_A \min_{\bar\varepsilon} \left\{2\delta^{-1}(A\bar\varepsilon, p) + (MA\bar\varepsilon, A\bar\varepsilon)\right\},$$

the maximum being over $m \times n$ matrices A whose columns all have ℓ^2 norm at most 1, and the minimum is over $\bar\varepsilon \in \{-1,1\}^n$. In Lemma 5.4, we describe the behavior of $H_{\delta,n}(M,p)$ as $\delta \to 0$ using the language of half-relaxed limits. We show that when $p \neq 0$

$$\lim_{\delta\to 0} H_{\delta,n}(M,p) = \max_A \min_{\bar\varepsilon} \left\{(\Pi(p)M\Pi(p)A\bar\varepsilon, A\bar\varepsilon)\right\},$$

where $\Pi(p)$ is the $m \times m$ matrix representing the orthogonal projection onto the hyperplane orthogonal to p. (We leave the discussion of the case when $p = 0$ to Section 5.) This is where the convergence proof departs most from [78], and this is due to the different set of admissible matrices A. We take advantage of Banaszczyk's theorem (Theorem 5.1) to characterize this last maxmin, obtaining (see Lemma 5.6) that (still for $p \neq 0$)

$$\max_A \min_{\bar\varepsilon} \left\{(\Pi(p)M\Pi(p)A\bar\varepsilon, A\bar\varepsilon)\right\} = \sum_{i=0}^{k-1} \sigma_{m-i}((\Pi(p)M\Pi(p))_+) =: H_n(M,p),$$

where $k = \min(n, m-1)$ and for a given symmetric matrix M we use M_+ to denote the positive semi-definite part of M, and $\sigma_1(M), \dots, \sigma_m(M)$ to denote the eigenvalues of M ordered from smallest to largest (counting multiplicities).

This argument enables one to identify what the limiting PDE should be, it is akin to checking the consistency of a numerical scheme, except here we are finding the PDE itself as we pass to the limit. In concrete terms, the above shows that if $\phi = \phi(x,t)$ is a smooth function (say twice differentiable in x, differentiable in t) then the limit

$$\lim_{\delta\to 0} \left\{ \tfrac{1}{\delta^2}(\phi(x,t) - \phi(x,t-\delta^2)) + \max_A \min_{\bar\varepsilon} \left\{\delta^{-2}(\phi(x+\delta A\bar\varepsilon, t) - \phi(x,t))\right\} \right\},$$

exists and is equal to (provided $\nabla\phi(x,t) \neq 0$)

$$\partial_t \phi(x,t) + \tfrac{1}{2} H_n(D^2\phi, \nabla\phi)(x,t).$$

The finite difference operator in the equation solved by w_δ converges to the above operator as $\delta \to 0$. As explained at the end of Section 5 this corresponds to the level set PDE for a curvature flow (and, in particular, to the mean curvature flow if $n \geq m-1$).

Here is where the theory of viscosity solutions comes in (see Appendix B for definitions). This theory is used in Section 6 where the proof of convergence for the value function is finalized for E is smooth (Theorem 6.1). Viscosity solutions are based on the comparison principle and the monotonicity of the differential operators involved; it is for this reason this notion of solution enjoys strong stability properties. In fact, the stability is strong enough to preserve being a solution to the finite difference equation (3.2) in the $\delta \to 0$ limit without requiring any a priori compactness of the w_δ (beyond boundedness). Loosely speaking, this is achieved by showing that every potential limit point of the sequence $w_\delta(x,t)$ lies above a viscosity supersolution and below a viscosity subsolution to the limiting PDE described earlier (see Lemma 6.5). The comparison principle says that the subsolution must lie below the supersolution, and therefore they must be one and the same function, guaranteeing the whole sequence w_δ converges to a limit (see Remark 6.3).

The proof of Theorem 2.1 is completed in the last part of Section 6 by first using Theorem 6.1 to verify the convergence of $\mathcal{K}_T(E)$ for smooth E. This requires relating the value of $w(0,-1)$ (for w solving the limiting PDE) to the extinction time for the corresponding curvature flow starting from E. The general case is done via an approximation, making use of the fact that given a non-smooth E we can find for any given $\epsilon > 0$ a centrally symmetric smooth convex body E_ϵ such that

$$(1-\epsilon)E_\epsilon \subset E \subset (1+\epsilon)E_\epsilon.$$

From here, we are able to estimate $\mathcal{K}_T(E)$ in terms of $\mathcal{K}_T(E_\epsilon)$ in a uniform way in T, which allow us to conclude the proof for a general (centrally symmetric, convex) E, including most importantly, $E = [-1,1]^m$.

Overview of Theorem 2.2 on estimates for the extinction time. This is the more general counterpart to Theorem 1.2. In the discussion on Theorem 2.1, we saw the relationship between $\tau_n(E)$ and $w(0,-1)$, so the two-sided bound for $\tau_n(E)$ are obtained by obtaining a respective two-sided bound for $w(0,-1)$. As w solves a parabolic PDE, one may use the comparison principle (see Theorem B.2) together with adequately built *barriers* to estimate w from above or below. For simplicity of exposition, we will only discuss here the plan of the proof for $E = [-1,1]^m$.

Let us review how the comparison principle is used. As we know, w solves (in the viscosity sense)

$$\partial_t w + \tfrac{1}{2} H_n(D^2 w, \nabla w) = 0, \quad \text{in } \mathbb{R}^m \times \mathbb{R}_-.$$

If a continuous function ψ is such that $\psi(x,0) \le w(x,0)$ and satisfies the differential inequality

$$\partial_t \psi + \tfrac{1}{2} H_n(D^2 \psi, \nabla \psi) \ge 0, \quad \text{in } \mathbb{R}^m \times \mathbb{R}_-,$$

then for all x and t we have $\psi(x,t) \le w(x,t)$. The same statement holds if we reverse the inequalities, i.e. if we have a continuous function $\overline{\psi}$ such that $\overline{\psi}(x,0) \ge w(x,0)$ and

$$\partial_t \overline{\psi} + \tfrac{1}{2} H_n(D^2 \overline{\psi}, \nabla \overline{\psi}) \le 0, \quad \text{in } \mathbb{R}^m \times \mathbb{R}_-.$$

Then it follows that $\overline{\psi}(x,t) \ge w(x,t)$ for all x,t. In the PDE nomenclature one says ψ is a subsolution and $\overline{\psi}$ is a supersolution (see Appendix B for exact definitions). An auxiliary (often explicit) function that is used as a subsolution or a supersolution via the comparison principle is colloquially known as a *barrier*.

In the proof of Theorem 2.2, we proceed by constructing functions $\psi, \overline{\psi}$ which are respectively a subsolution and a supersolution to the PDE solved by w. Then, we apply the comparison principle twice to conclude that $\psi \le w \le \overline{\psi}$ everywhere. This yields a two sided estimate for $w(0,-1)$

$$\psi(0,-1) \le w(0,-1) \le \overline{\psi}(0,-1).$$

The functions $\psi, \overline{\psi}$ we use turn out to be solutions to the heat equation (with different diffusivity constants, and different terminal values). In particular, the values $\psi(0,-1)$ and $\overline{\psi}(0,-1)$ correspond to, respectively, the expected largest component and the ℓ_∞ norm of Gaussians (of different variance), yielding the bounds in the statement of the Theorem.

It remains to explain how the heat equation is used to produce a subsolution ψ and a supersolution $\overline{\psi}$ for the nonlinear PDE for w. First, as we intend ψ to bound w from below, we choose as its terminal value the function $x \mapsto \max_i x_i$ which lies below $\|x\|_\infty$. So, we define ψ as the unique solution to

$$\partial_t \psi + \tfrac{1}{4} \Delta \psi = 0 \text{ for } t < 0, \ \psi(x,0) = \max_i x_i.$$

The second derivatives of the function ψ satisfy the following for $t < 0$,

$$\partial_{ij} \psi \le 0 \text{ whenever } i \ne j, \ \text{ and } \sum_{j=1}^m \partial_{ij} \psi = 0 \text{ for every } i.$$

This means that $-D^2\psi$ is a Laplacian matrix.[6] With this, it is possible to show the function ψ satisfies the bound (see Lemma 7.3)

$$\partial_t \psi + \tfrac{1}{2} H_n(D^2 \psi, \nabla \psi) \ge 0.$$

We see that ψ is a subsolution, as we wanted.

Now, we define $\overline{\psi}$ as the solution of (note the different diffusivity constant)

$$\partial_t \overline{\psi} + \tfrac{1}{2} \Delta \overline{\psi} = 0 \text{ for } t < 0, \ \psi(x,0) = \|x\|_\infty.$$

[6] We are using here the convention that a Laplacian matrix is negative semi-definite for consistency with the standard Laplacian operator in $\mathbb{R}^m$. However, in parts of the literature the sign convention in the definition of the Laplacian matrix is the opposite to ours, making it a positive semi-definite matrix.

Observe $\psi(x,0)$ is convex in x, thus a standard computation with the heat kernel shows that $D^2\overline{\psi} \geq 0$ for all $t<0$. In particular, whenever $\nabla\overline{\psi} \neq 0$ we have

$$H_n(D^2\overline{\psi}, \nabla\overline{\psi}) = \mathrm{tr}(\Pi(\nabla\overline{\psi})D^2\overline{\psi}) \leq \mathrm{tr}(D^2\overline{\psi}) = \Delta\overline{\psi},$$

(recall we are in the case where $n \geq m-1$), and so

$$\partial_t\overline{\psi} + \tfrac{1}{2}H_n(D^2\overline{\psi}, \nabla\overline{\psi}) \leq \partial_t\overline{\psi} + \frac{1}{2}\Delta\overline{\psi} = 0,$$

and we see $\overline{\psi}$ is a supersolution for the PDE solved by w, as we wanted.

4. The value function

In all that follows, we fix a dimension $m \geq 2$ and a batch size $n \geq 1$. Then, for any centrally symmetric convex body E the iterative maxmin formulation the value function v given by Definition 3.1 allows us to characterize it by a dynamic program:

$$\begin{cases} v(y,s-1) = & \max_{A}\min_{\bar{\varepsilon}} v(y+A\bar{\varepsilon},s) \quad \text{for } s \leq 0, \\ \quad v(y,0) = & \|y\|_E. \end{cases} \tag{4.1}$$

The maximum is over the set $\mathcal{A}$ of all $m\times n$ matrices whose n columns $a_1,\dots,a_n$ are all such that $|a_i| \leq 1$, and the minimum is over $\bar{\varepsilon} \in \{-1,1\}^n$.

Remark 4.1. *When $E=[-1,1]^m$, it is clear that*

$$v(0,-1) = \max_{A}\min_{\bar{\varepsilon}} \|A\bar{\varepsilon}\|_\infty = K_1(m,n).$$

Remark 4.2. *The function v characterizes the value of $\mathcal{K}_{n,T}(E)$, to be precise,*

$$\mathcal{K}_{n,T}(E) = v(0,-T).$$

Accordingly, the analysis of the long-time asymptotics of $\mathcal{K}_{n,T}(E)$ involves understanding v as $s \to -\infty$.

To start our analysis, we gather some basic properties of v.

Proposition 4.3. *Given any centrally symmetric convex body E and v as in (4.1), we have*

(1) *For every s the function $v(\cdot,s)$ is 1-Lipschitz in y with respect to the $\|\cdot\|_E$ norm.*
(2) *The value function is monotone in s, namely*

$$v(y,s-1) \geq v(y,s) \text{ for all } (y,s) \in \mathbb{R}^m\times\mathbb{Z}_-.$$

Proof. We argue by induction on the non-negative integer $-s$. For the first claim, we note that the fact that the function $\|y\|_E$ is 1-Lipschitz with respect to $\|\cdot\|_E$ follows from the triangle inequality for the norm $\|\cdot\|_E$. This proves the claim for $s=0$.

Now suppose that for some $s<0$ we know that $v(y,s)$ is 1-Lipschitz continuous in y, that is

$$v(y,s) \leq v(y+h,s) + \|h\|_E \text{ for all } y,h\in\mathbb{R}^m.$$

Then, in particular, for any matrix A, any $\bar{\varepsilon}\in\{-1,1\}^n$, and $y,h\in\mathbb{R}^m$ we have

$$v(y+A\bar{\varepsilon},s) \leq v(y+A\bar{\varepsilon}+h,s) + \|h\|_E \text{ for all } y,h\in\mathbb{R}^m.$$

If we fix one such A and consider the minimum in $\bar{\varepsilon}$ we have

$$\min_{\bar{\varepsilon}}\{v(y+A\bar{\varepsilon},s)\} \leq \min_{\bar{\varepsilon}}\{v(y+A\bar{\varepsilon}+h,s)\} + \|h\|_E \text{ for all } y,h\in\mathbb{R}^m.$$

Then, taking the maximum of the result over all A

$$\max_{A}\min_{\bar{\varepsilon}}\{v(y+A\bar{\varepsilon},s)\} \leq \max_{A}\min_{\bar{\varepsilon}}\{v(y+A\bar{\varepsilon}+h,s)\} + \|h\|_E \text{ for all } y,h\in\mathbb{R}^m.$$

In light of the construction of the function v, this last inequality can be rewritten as

$$v(y,s-1) \leq v(y+h,s-1) + \|h\|_E \text{ for all } y,h\in\mathbb{R}^m.$$

We conclude that $v(\cdot,s-1)$ is also 1-Lipschitz with respect to $\|\cdot\|_E$, and the first claim is proved.

As for the second claim, simply note that for each s, the matrix $A=0$ is admissible, so $v(y,s-1)\geq v(y,s)$. □

The following pointwise bound for $v(y,s)$ will be important, it follows via a comparison argument.

Lemma 4.4. *If E has a C^2 boundary, there exists a constant $C \equiv C(E,n)$ such that*

$$v(y,s) \le \sqrt{\|y\|_E^2 + C|s|} \text{ for all } (y,s) \in \mathbb{R}^m \times \mathbb{Z}_-.$$

Proof. For a number $C > 0$ to be determined below, we consider the function

$$b(y,s) = \sqrt{\|y\|_E^2 - Cs} \text{ defined for all } (y,s) \in \mathbb{R}^m \times \mathbb{Z}_-.$$

The lemma will follow from the fact that b is a supersolution for the discrete-in-time terminal value problem that characterizes v. Specifically, we will show that

$$\begin{aligned} b(y,s) &\ge \max_A \min_{\bar{\varepsilon}} b(y + A\bar{\varepsilon}, s+1) \ \forall\, y \in \mathbb{R}^m, s < 0, \\ b(y,0) &= \|y\|_E \ \forall\, y \in \mathbb{R}^m \end{aligned}$$

The last equation is straightforward: regardless of the value of C, we have that $b(y,0) = \|y\|_E$. The first inequality requires comparing $\|y + A\varepsilon\|_E^2$ to $\|y\|_E^2$ for every admissible A and a corresponding $\bar{\varepsilon}$. To prove this inequality, note that since $b(y,s)$ is always non-negative, it is sufficient to show that for all $y \in \mathbb{R}^m$ and all $s < 0$,

$$b(y,s)^2 \ge \max_A \min_{\bar{\varepsilon}} b(y + A\bar{\varepsilon}, s+1)^2,$$

or more explicitly, we only need to show that

$$\|y\|_E^2 - Cs \ge \max_A \min_{\bar{\varepsilon}} \left\{ \|y + A\bar{\varepsilon}\|_E^2 - C(s+1) \right\}.$$

This in turn is equivalent to

$$\|y\|_E^2 + C \ge \max_A \min_{\bar{\varepsilon}} \left\{ \|y + A\bar{\varepsilon}\|_E^2 \right\}.$$

To establish this last inequality, we make the following standard observation:

$$\min_{\bar{\varepsilon}} \|y + A\bar{\varepsilon}\|_E^2 \le \min_{\bar{\varepsilon}} \left\{ \tfrac{1}{2}\|y + A\bar{\varepsilon}\|_E^2 + \tfrac{1}{2}\|y - A\bar{\varepsilon}\|_E^2 \right\}.$$

This means that

$$\max_A \min_{\bar{\varepsilon}} \|y + A\bar{\varepsilon}\|_E^2 \le \max_A \min_{\bar{\varepsilon}} \left\{ \tfrac{1}{2}\|y + A\bar{\varepsilon}\|_E^2 + \tfrac{1}{2}\|y - A\bar{\varepsilon}\|_E^2 \right\}.$$

Therefore, it suffices to show that there is $C > 0$ large enough so that

$$\max_A \min_{\bar{\varepsilon}} \left\{ \tfrac{1}{2}\|y + A\bar{\varepsilon}\|_E^2 + \tfrac{1}{2}\|y - A\bar{\varepsilon}\|_E^2 \right\} \le \|y\|_E^2 + C, \ \forall\, y \in \mathbb{R}^m.$$

Rearranging this expression, we get

$$\max_A \min_{\bar{\varepsilon}} \left\{ \tfrac{1}{2}\|y + A\bar{\varepsilon}\|_E^2 + \tfrac{1}{2}\|y - A\bar{\varepsilon}\|_E^2 - \|y\|_E^2 \right\} \le C, \ \forall\, y \in \mathbb{R}^m.$$

Since the boundary of E is C^2, the function $y \mapsto \|y\|_E^2$ is at least locally $C^{1,1}$ and homogeneous of degree 2, and its Hessian exists a.e. and is a 0-homogeneous function in $\mathbb{R}^m$. Therefore, the operator norm of the Hessian is bounded uniformly on $\mathbb{R}^m$ by some constant $L(E) > 0$ a.e. For the function $g(z) := \frac{1}{2}\|y + zA\bar{\varepsilon}\|_E^2$, the optimization objective above can be bounded as follows:

$$g(1) + g(-1) - 2g(0) = \int_0^1 g'(z) - g'(z-1)dz \le \|g''(z)\|_{L_\infty([-1,1])} \le \frac{1}{2}L(E)\|A\bar{\varepsilon}\|_2^2 \le C(E,n)$$

where $C(E,n) := \frac{1}{2}L(E)n^2$. □

Remark 4.5. *We will not use the exact value $C(E,n)$ in our estimate of v. Instead, in our proofs, we will only use the fact that $C(E,n)$ is a finite number to get uniform pointwise bounds on v as we proceed to the continuum limit.*

5. Banaszczyk's theorem and the limiting Bellman operator

We introduce a scaling parameter and proceed towards a continuum limit, following Kohn and Serfaty [78]. Working with a parabolic scaling with spatial scale δ we define for $x \in \mathbb{R}^m$ and $t \in \delta^2\mathbb{Z}_-$

$$w_\delta(x,t) := \delta v(\delta^{-1}x, \delta^{-2}t).$$

The value function $v(y,s)$ solves (4.1), so w_δ will solve a rescaled version. Observe that

$$\begin{aligned} w_\delta(x,t-\delta^2) &= \delta v(\delta^{-1}x, \delta^{-2}(t-\delta^2)) \\ &= \max_A \min_{\bar\varepsilon} \delta v(\delta^{-1}x + A\bar\varepsilon, \delta^{-2}(t-\delta^2)+1) \\ &= \max_A \min_{\bar\varepsilon} \delta v(\delta^{-1}(x+\delta A\bar\varepsilon), \delta^{-2}t). \end{aligned}$$

Moreover, for every x we have $w_\delta(x,0) = \delta\|x/\delta\|_E = \|x\|_E$. The above shows that w_δ solves

$$w_\delta(x,t-\delta^2) = \max_A \min_{\bar\varepsilon} w_\delta(x+\delta A\bar\varepsilon, t) \quad \text{for } t \in \delta^2\mathbb{Z}_-, \tag{5.1}$$

$$w_\delta(x,0) = \|x\|_E. \tag{5.2}$$

The following theorem of Banaszczyk [9, Theorem 1] will be key for this section; it allows us to characterize the Bellman equation that arises in the limit $\delta \to 0$.

Theorem 5.1 (Banaszczyk's theorem)**.** *Let D be an ellipsoid in $\mathbb{R}^k$ with center at 0 and principal semi-axes $\lambda_1, \ldots, \lambda_k$. Then, for any finite set of vectors $z_1, \ldots, z_n \in D$ we can find signs $\bar\varepsilon \in \{\pm 1\}^n$ such that*

$$|\varepsilon_1 z_1 + \ldots + \varepsilon_n z_n|^2 \leq \sum_{j=1}^k \lambda_j^2.$$

In the following remark we rephrase this theorem in equivalent terms that will make its connection to the Bellman equation more apparent.

Remark 5.2. *If $M \in \mathbb{R}^{m\times m}$ is positive semi-definite and $k = \min(n, rank(M))$, then*

$$\max_A \min_{\bar\varepsilon} (MA\bar\varepsilon, A\bar\varepsilon) = \sum_{j=0}^{k-1} \sigma_{m-j}(M), \tag{5.3}$$

where the feasible sets of A and $\bar\varepsilon$ are as usual.

To see why this remark holds, consider an admissible matrix A whose columns are denoted $a_1, \ldots, a_n$ and write $z_i = M^{1/2}a_i$ for each i. Then we observe that

$$(MA\bar\varepsilon, A\bar\varepsilon) = (M(\varepsilon a_1 + \ldots + \varepsilon_n a_n), \varepsilon a_1 + \ldots + \varepsilon_n a_n) = |\varepsilon_1 z_1 + \ldots + \varepsilon_n z_n|^2.$$

If the a_i's lie in some k-dimensional subspace V, and $Ma_i \neq 0$ for every i, then the z_i will all lie in an ellipsoid with squares of principal semi axes adding up to no more than the partial trace $\mathrm{tr}_V(M)$. Then, Theorem 5.1 guarantees that

$$\min_{\bar\varepsilon} (MA\bar\varepsilon, A\bar\varepsilon) \leq \mathrm{tr}_V(M).$$

We may now optimize over all such subspaces k, and conclude that

$$\max_A \min_{\bar\varepsilon} (MA\bar\varepsilon, A\bar\varepsilon) \leq \max_{\dim(V)=k} \mathrm{tr}_V(M).$$

For a symmetric matrix M the maximum value of $\mathrm{tr}_V(M)$ over all k-dimensional subspaces is the same as the sum of the largest k eigenvalues of M (each again value counted according to its multiplicity), so

$$\max_A \min_{\bar\varepsilon} (MA\bar\varepsilon, A\bar\varepsilon) \leq \sum_{j=0}^{k-1} \sigma_{m-j}(M).$$

On the other hand, if A_0 is chosen with columns given by an eigen-basis for M for the largest k eigenvalues of M (and zero for the remaining columns if $n > k$), we see that

$$(MA_0\bar\varepsilon, A_0\bar\varepsilon) = \sum_{j=0}^{k-1} \sigma_{m-j}(M), \ \forall\, \bar\varepsilon \in \{-1,1\}^n,$$

and the identity in (5.3) now follows.

5.1. **The difference operators.** Subtracting $w_\delta(x,t)$ from both sides of (5.1) gives us

$$w_\delta(x,t) - w_\delta(x,t-\delta^2) + \max_A \min_{\bar{\varepsilon}} \{w_\delta(x+\delta A\bar{\varepsilon},t) - w_\delta(x,t)\} = 0,$$

for every $x \in \mathbb{R}^m$, $t \in \delta^2\mathbb{Z}_-$. We shall rewrite this equation in terms of the following finite difference operators, which act on functions $\phi(x,t)$ defined in $\mathbb{R}^m \times \delta^2\mathbb{Z}_-$. They are

$$\partial_t^{-\delta^2}\phi(x,t) := \frac{1}{\delta^2}(\phi(x,t) - \phi(x,t-\delta^2))$$
$$\mathcal{H}_{\delta,n}(\phi)(x,t) := \max_A \min_{\bar{\varepsilon}} \tfrac{2}{\delta^2}(\phi(x+\delta A\bar{\varepsilon},t) - \phi(x,t))$$

In terms of these operators, the equation for w_δ can be written more compactly,

$$\begin{cases} \partial_t^{-\delta^2} w_\delta(x,t) + \frac{1}{2}\mathcal{H}_{\delta,n}(w_\delta)(x,t) &= 0 \text{ in } \mathbb{R}^m \times \delta^2\mathbb{Z}_-, \\ w_\delta(x,0) &= \|x\|_E \text{ in } \mathbb{R}^m. \end{cases} \tag{5.4}$$

Evidently, as $\delta \to 0$ one expects the linear difference operator $\partial_t^{-\delta^2}$ to converge to the ∂_t operator. Similarly, we expect the nonlinear operator $\mathcal{H}_{\delta,n}$ to converge to a nonlinear elliptic operator. For a matrix $M \in \mathbb{R}^{m\times m}$ and $p \in \mathbb{R}^m$ we define $H_{\delta,n}(M,p)$ as follows:

$$H_{\delta,n}(M,p) := \max_A \min_{\bar{\varepsilon}} \left\{2\delta^{-1}(p, A\bar{\varepsilon}) + (MA\bar{\varepsilon}, A\bar{\varepsilon})\right\}. \tag{5.5}$$

Proposition 5.3. *Consider a function $\phi : \mathbb{R}^m \times (-\infty, 0) \to \mathbb{R}$ such that $\partial_t\phi$ and $D^2\phi$ are continuous with modulus of continuity $\rho(\cdot)$, then the following estimates hold for all (x,t)*

$$|\partial_t^{-\delta^2}\phi(x,t) - \partial_t\phi(x,t)| \le \rho(\delta^2), \;\; |\mathcal{H}_{\delta,n}(\phi)(x) - H_{\delta,n}(D^2\phi(x), \nabla\phi(x))| \le 2\rho(\delta n)n^2.$$

The proof of this proposition is standard and it can be found in Appendix A. The operator $H_{\delta,n}(M,p)$ captures how $\mathcal{H}_{\delta,n}$ is converging from a discrete (and thus, nonlocal) operator to a differential (local) operator as $\delta \to 0$. While Proposition 5.3 deals with a fixed function ϕ (namely, independent of δ) and not with the sequence w_δ, we will be able to relate (in a weak sense) the limiting behavior of the two sequences

$$\{\mathcal{H}_\delta(w_\delta)(x,t)\}_{\delta>0} \text{ and } \{H_{\delta,n}(D^2 w_\delta, \nabla w_\delta)(x,t)\}_{\delta>0}.$$

In this way, the sequence $H_{\delta,n}$ captures all of the relevant information of the operator $\mathcal{H}_{\delta,n}$ as $\delta \to 0$. The fact that $H_{\delta,n}$ is a sequence of real valued functions in a finite dimensional space (that is the Cartesian product of the space of symmetric $m \times m$ matrices and $\mathbb{R}^m$) reflects the emergence of a nonlinear local operator of the form $\phi \mapsto H(D^2\phi, \nabla u)$ in the $\delta \to 0$ limit.

To describe the limit of $H_{\delta,n}$ it will be useful to define a subset of our family of admissible matrices for each given $p \in \mathbb{R}^m$:

$$\mathcal{A}(p) := \{A = (a_1 \;\ldots\; a_n) : |a_i| \le 1 \text{ and } (a_i, p) = 0 \text{ for all } i\} = \{A \in \mathcal{A} \;:\; A^\top p = 0\}. \tag{5.6}$$

We then define $H_n : \mathrm{Sym}(m) \times \mathbb{R}^m \to \mathbb{R}$ by

$$H_n(M,p) = \max_{A\in\mathcal{A}(p)} \min_{\bar{\varepsilon}\in\{\pm1\}^n} (MA\bar{\varepsilon}, A\bar{\varepsilon}) \tag{5.7}$$

The function $H_n(M,p)$ is finite and well-defined for every M and p since the expression inside the $A \mapsto M(A\bar{\varepsilon}, A\bar{\varepsilon})$ is continuous with respect to A and $\mathcal{A}(p)$ is always a compact set. Moreover, $H_n(M,p)$ is continuous in the set $\{(M,p) : p \neq 0\}$.

On the other hand, H_n cannot be extended to a continuous function for all (M,p), i.e. there are discontinuities at some (M,p)'s with $p = 0$. This is to be expected, as for $p \neq 0$ the function $p \mapsto H_n(M,p)$ depends on p only on the direction of p. If the eigenvalues of M are all different, a sequence p_k with $p_k \to 0$ along several directions will produce values $H_n(M,p_k)$ that oscillate indefinitely between different values.

However, we can and must work with the upper and lower semi-continuous envelopes of H_n given by

$$(H_n)^*(M,p) := \limsup_{M'\to M, p'\to p} H_n(M',p'),$$
$$(H_n)_*(M,p) := \liminf_{M'\to M, p'\to p} H_n(M',p').$$

These envelopes will be needed to define the notion of weak solution for the limiting PDE. This is a well known practice in the viscosity solutions literature, in particular works on mean curvature flows [41, 49].

We already noted that H_n is continuous when $p \neq 0$, in this case $H_n^*(M,p) = H_{n,*}(M,p)$ for $p \neq 0$. One can show that

$$(H_n)^*(M,0) = \max_A \min_{\overline{\varepsilon}} (MA\overline{\varepsilon}, A\overline{\varepsilon}),$$
$$(H_n)_*(M,0) = \min_{\sigma \in \mathbb{S}^{m-1}} H_n(M,\sigma).$$

We recall the maximum in the first line is over all A's whose columns have ℓ^2 norm at most 1, or in the notation just introduced, $A \in \mathcal{A}(0)$. In particular, this shows that $(H_n)^*(M,0) = H_n(M,0)$ so while H_n is not continuous for all M,p it is always upper semi-continuous. We will establish this in the following proposition, where we do something stronger: we characterize the half-relaxed limits of $H_\delta(M,p)$ as $\delta \to 0$ (see Remark 5.5 below for further details).

Lemma 5.4. *For any sequences*

$$M_k \to M,\ p_k \to p,\ \delta_k \to 0,$$

of symmetric matrices $M_k \in \mathbb{R}^{m\times m}$, vectors $p_k \in \mathbb{R}^m$ and scalars $\delta_k > 0$, we have the limit

$$\lim_{k\to\infty} H_{\delta_k,n}(M_k,p_k) = H_n(M,p), \quad \textit{when } p \neq 0,$$

as well as the bounds

$$\liminf_{k\to\infty} H_{\delta_k,n}(M_k,p_k) \geq (H_n)_*(M,0) \quad \textit{and} \ \limsup_{k\to\infty} H_{\delta_k,n}(M_k,p_k) \leq (H_n)^*(M,0) \ \textit{when } p = 0.$$

Moreover, both bounds are sharp, i.e., for both bounds one can find sequences where the bounds hold with equality.

Remark 5.5. *Another way of stating Lemma 5.4 is in terms of the half-relaxed limits of $H_{\delta,n}$ as $\delta \to 0$. In particular, the lemma says that when $p = 0$,*

$$(H_n)^*(M,p) = \limsup_{M'\to M, p'\to p, \delta\to 0} H_{\delta,n}(M',p') = \max_A \min_{\overline{\varepsilon}} (MA\overline{\varepsilon}, A\overline{\varepsilon}),$$
$$(H_n)_*(M,p) = \liminf_{M'\to M, p'\to p, \delta\to 0} H_{\delta,n}(M',p') = \min_{\sigma\in\mathbb{S}^{m-1}} H_n(M,\sigma),$$

and furthermore, $(H_n)^*(M,p) = (H_n)_*(M,p) = H_n(M,p)$ whenever $p \neq 0$.

Proof. We start by fixing an arbitrary $\delta > 0$ and recording two observations about $H_{\delta,n}$. First, it is clear that restricting the maximization in A to just those A's with $A \in \mathcal{A}(p)$ cannot increase the maximum, and so we have for any symmetric M and vector p,

$$H_{\delta,n}(M,p) \geq \max_{A\in\mathcal{A}(p)} \min_{\overline{\varepsilon}} (MA\overline{\varepsilon}, A\overline{\varepsilon}) = H_n(M,p). \tag{5.8}$$

Here, we have made use of the fact that $A^\top p = 0$ whenever $A \in \mathcal{A}(p)$.

The second observation concerns $H_{\delta,n}$ when $p \neq 0$, we are going to show that given any symmetric M and a vector $p \neq 0$, if A achieves the maximum in (5.5) we have

$$\|(I - \Pi(p))A\| = \|\hat{p} \otimes \hat{p}A\| \leq \tfrac{1}{2}\delta|p|^{-1}C(m,n)\|M\|, \tag{5.9}$$

where $\Pi(p) := I - \hat{p}\otimes\hat{p}$, $\hat{p} := p/|p|$ for $p \neq 0$. To prove (5.9) we first note that the minimization in (5.5) involves an objective function that is the sum of an odd and an even function of $\overline{\varepsilon}$. Keeping this in mind, we compare the values of the objective function at $\overline{\varepsilon}$ and $-\overline{\varepsilon}$, and conclude that

$$\min_{\overline{\varepsilon}}\{2\delta^{-1}(p, A\overline{\varepsilon}) + (MA\overline{\varepsilon}, A\overline{\varepsilon})\} = \min_{\overline{\varepsilon}}\{-2\delta^{-1}|(p, A\overline{\varepsilon})| + (MA\overline{\varepsilon}, A\overline{\varepsilon})\}. \tag{5.10}$$

Then, writing $A = \Pi(p)A + (I - \Pi(p))A$, where ,

$$(p, A\overline{\varepsilon}) = (p, (I-\Pi(p))A\overline{\varepsilon}) = |p|(A^\top\hat{p}, \overline{\varepsilon}).$$

Then, given that A achieves the maximum in (5.5) for a given M and p,

$$0 \leq \min_{\overline{\varepsilon}}\{-2\delta^{-1}|(p, A\overline{\varepsilon})| + (MA\overline{\varepsilon}, A\overline{\varepsilon})\},$$

and thus for every $\overline{\varepsilon}$ we have

$$2\delta^{-1}|(A^\top p, \overline{\varepsilon})| \leq (MA\overline{\varepsilon}, A\overline{\varepsilon}) \leq C(m,n)\|M\|.$$

Then, choosing $\overline{\varepsilon}$ for which

$$(A^\top p, \overline{\varepsilon}) = |(Ae_1)^\top \cdot p| + \ldots + |(Ae_n)^\top \cdot p| = \|A^\top p\|_1,$$

we conclude that

$$\|A^\top p\|_1 \leq \tfrac{1}{2}\delta C(m,n)\|M\|.$$

Then, noting that $\|\hat{p} \otimes \hat{p}A\| = |A^\top \hat{p}| \leq |p|^{-1}\|A^\top p\|_1$ we obtain (5.9).

With these basic estimates in hand, consider a generic sequence M_k, p_k, δ_k as in the statement of the lemma. First, take the case where $p_k \to p \neq 0$. If we let A_k denote a matrix achieving the maximum for (M_k, p_k), estimate (5.9) says (without loss of generality, $p_k \neq 0$ for every k)

$$\|(I - \Pi(p_k))A_k\| \leq \tfrac{1}{2}\delta_k |p_k|^{-1} C(m,n)\|M_k\| \;\; \forall\, k.$$

On account of $p \neq 0$ we know that for all sufficiently large k we have $\frac{1}{2}|p_k|^{-1} \leq |p|^{-1}$. Likewise, for large enough k we have $\|M_k\| \leq 1 + 2\|M\|$. Therefore for all large enough k we have the inequality

$$\|(I - \Pi(p_k))A_k\| \leq \delta_k |p|^{-1} C(m,n)(1 + 2\|M\|).$$

We use this to estimate $(M_k A_k \overline{\varepsilon}, A_k \overline{\varepsilon})$ for large k, via the decomposition $A_k = \Pi(p_k)A_k + (I - \Pi(p_k))A_k$,

$$(M_k A_k \overline{\varepsilon}, A_k \overline{\varepsilon}) \leq (M_k \Pi(p_k) A_k \overline{\varepsilon}, \Pi(p_k) A_k \overline{\varepsilon}) + C\|M_k\|\|(I - \Pi(p_k))A_k\|.$$

Then for all large enough k and for every $\overline{\varepsilon}$ we have

$$(M_k A_k \overline{\varepsilon}, A_k \overline{\varepsilon}) \leq \{(M_k \Pi(p_k) A_k \overline{\varepsilon}, \Pi(p_k) A_k \overline{\varepsilon})\} + C(m,n)(1 + 2\|M\|)^2 |p|^{-1}\delta_k.$$

From here it follows that

$$\begin{aligned} H_{\delta_k,n}(M_k, p_k) &\leq \min_{\overline{\varepsilon}} (M_k(\Pi(p_k)A_k)\overline{\varepsilon}, (\Pi(p_k)A_k)\overline{\varepsilon}) + C(m,n)(1 + 2\|M\|)^2 |p|^{-1}\delta_k \\ &\leq H_n(M_k, p_k) + C(m,n)(1 + 2\|M\|)^2 |p|^{-1}\delta_k, \end{aligned}$$

where we used that $\Pi(p_k)A_k \in \mathcal{A}(p_k)$ to obtain the second inequality. Passing to the limit, we have

$$\limsup_{k \to \infty} H_{\delta_k,n}(M_k, p_k) \leq H_n(M, p).$$

The case $p \neq 0$ now follows simply by applying (5.8) to $H_{\delta_k}(M_k, p_k)$, which leads to

$$\liminf_{k \to \infty} H_{\delta_k,n}(M_k, p_k) \geq \lim_{k \to \infty} H_n(M_k, p_k) = H_n(M, p).$$

This shows $(H_n)^*(M,p) = (H_n)_*(M,p) = H_n(M,p)$ for $p \neq 0$.

Now we consider what happens when $p_k \to 0$. For the liminf bound, we use equation (5.8) once again:

$$H_{\delta,n}(M,p) \geq H_n(M,p) \geq \min_q H_n(M,q).$$

The minimum being over all $q \in \mathbb{R}^m$, noting that the minimum cannot be achieved only at $q = 0$, and using that $H_n(M,q)$ is 0-homogeneous for $q \neq 0$, we obtain in fact that

$$H_{\delta,n}(M,p) \geq \min_{\sigma \in \mathbb{S}^{m-1}} H_n(M,\sigma).$$

Since this is true for every M, p, δ, it follows that

$$\liminf_{k \to \infty} H_{\delta_k,n}(M_k, p_k) \geq \min_{\sigma \in \mathbb{S}^{m-1}} H_n(M,\sigma).$$

Now for the limsup bound, we use equation (5.10), which implies that

$$\max_A \min_{\overline{\varepsilon}} \{2\delta^{-1}(p, A\overline{\varepsilon}) + (MA\overline{\varepsilon}, A\overline{\varepsilon})\} \leq \max_A \min_{\overline{\varepsilon}} (MA\overline{\varepsilon}, A\overline{\varepsilon}) = H_n(M, 0).$$

Then, considering the sequence M_k, p_k, δ_k we have

$$\limsup_{k \to \infty} H_{\delta_k,n}(M_k, p_k) \leq H_n(M, 0),$$

which proves the bound for the limsup. To see the limsup upper bound is optimal, simply consider a sequence M_k, p_k, δ_k where $p_k = 0$ for every k. To see why the lower bound for the liminf is optimal, fix an arbitrary $\sigma_0 \in \mathbb{S}^{m-1}$ and consider a sequence with $p_k = (\delta_k)^{1/2}\sigma_0$, in this case, we have

$$\lim_{k\to\infty} H_{\delta_k,n}(M_k, p_k) = H_n(M, \sigma_0).$$

Indeed, this follows by applying (5.9), and noting that $\delta_k|p_k|^{-1} = \delta_k^{1/2}$, in which case the argument used for the case $p_k \to p \neq 0$ says that

$$\|(I - \Pi(p_k))A_k\| \leq \delta_k^{1/2}(1 + 2\|M\|)$$

and so the distance of A_k to $\mathcal{A}(\sigma_0)$ goes to zero as $k \to \infty$, proving the last limit holds. Since σ_0 is arbitrary, we can choose it to be the minimizer of $\sigma \mapsto H_n(M, \sigma)$, and this completes the proof.

□

We close this section with a more direct description of $H_n(M, p)$ when $p \neq 0$. We will use the decomposition of a given symmetric matrix M as $M = M_+ - M_-$ where M_+ and M_- are each positive semi-definite matrices with $M_+M_- = M_-M_+ = 0$. As before, for $p \neq 0$ we use $\Pi(p)$ to denote the orthogonal projection $I - |p|^{-2}p \otimes p$.

Lemma 5.6. *For $k = \min\{n, m-1\}$ and any symmetric $m \times m$ matrix M, we have*

$$H_n(M, p) = \sum_{j=0}^{k-1} \sigma_{m-j}((\Pi(p)M\Pi(p))_+), \quad \text{when } p \neq 0, \tag{5.11}$$

as well as $(H_n)^*(M, 0) = (H_n)^*(M_+, 0)$ and $(H_n)_*(M, 0) = \min_{\sigma\in\mathbb{S}^{m-1}} H_n((\Pi(\sigma)M\Pi(\sigma))_+, 0)$.

Proof. First, we note that as long as $p \neq 0$, then $A \in \mathcal{A}(p)$ if and only if $A = \Pi(p)A$ and $A \in \mathcal{A}(0)$. It follows that for any symmetric matrix M we have

$$\begin{aligned} H_n(M, p) &= \max_{A\in\mathcal{A}(p)} \min_{\overline{\varepsilon}}(MA\overline{\varepsilon}, A\overline{\varepsilon}) \\ &= \max_{A\in\mathcal{A}(p)} \min_{\overline{\varepsilon}}(M\Pi(p)A\overline{\varepsilon}, \Pi(p)A\overline{\varepsilon}) \\ &= \max_{A\in\mathcal{A}(0)} \min_{\overline{\varepsilon}}(\Pi(p)M\Pi(p)A\overline{\varepsilon}, A\overline{\varepsilon}). \end{aligned}$$

Put in succinct terms, this shows that when $p \neq 0$ we have

$$H_n(M, p) = H_n(\Pi(p)M\Pi(p), 0).$$

In light of this, we shall show that for any matrix $\tilde{M}$ we have

$$H_n(\tilde{M}, 0) = H_n(\tilde{M}_+, 0),$$

where $\tilde{M}_\pm$ are the aforementioned decomposition, that is $\tilde{M} = \tilde{M}_+ - \tilde{M}_-$.

Since $\tilde{M}_-$ is positive semi-definite, for any ξ we have

$$(\tilde{M}\xi, \xi) \leq (\tilde{M}_+\xi, \xi)$$

with equality holding if and only if $\tilde{M}_-\xi = 0$. Using that this holds for $\xi = A\overline{\varepsilon}$ for any A, ε it follows that

$$\max_{A\in\mathcal{A}(0)} \min_{\overline{\varepsilon}}(\tilde{M}A\overline{\varepsilon}, A\overline{\varepsilon}) \leq \max_{A\in\mathcal{A}(0)} \min_{\overline{\varepsilon}} \left\{(\tilde{M}_+A\overline{\varepsilon}, A\overline{\varepsilon})\right\}.$$

In other words, we have proved that

$$H_n(\tilde{M}, 0) \leq H_n(\tilde{M}_+, 0).$$

On the other hand, given any $A \in \mathcal{A}(0)$, there are matrices $A_1, A_2 \in \mathcal{A}(0)$ such that $A = A_1 + A_2$, $A_1, A_2 \in \mathcal{A}(0)$ and $\tilde{M}_-A_1 = 0$ and $\tilde{M}_+A_2 = 0$. In this case, we have

$$(\tilde{M}_+A\overline{\varepsilon}, A\overline{\varepsilon}) = (\tilde{M}A_1\overline{\varepsilon}, A_1\overline{\varepsilon})$$

and therefore,

$$\max_{A\in\mathcal{A}(0)} \min_{\overline{\varepsilon}}(\tilde{M}_+A\overline{\varepsilon}, A\overline{\varepsilon}) \leq \max_{A\in\mathcal{A}(0)} \min_{\overline{\varepsilon}}(\tilde{M}A\overline{\varepsilon}, A\overline{\varepsilon}).$$

This proves the reverse inequality,

$$H_n(\tilde{M}_+, 0) \leq H_n(\tilde{M}, 0).$$

We apply Theorem 5.1 with the matrix $\tilde{M}_+$ (using the observations made in Remark 5.2) and conclude that

$$H_n(\tilde{M}_+, 0) = \sum_{j=0}^{k-1} \sigma_{m-j}(\tilde{M}_+).$$

Applying the above to $\tilde{M} = \Pi(p)M\Pi(p)$ we complete the proof of the lemma in the case when $p \neq 0$.

As for $p = 0$, since $(H_n)^* = H_n$ the argument above already yields $(H_n)^*(M, 0) = (H_n)^*(M_+, 0)$. For $(H_n)_*$ we apply the arguments from $p \neq 0$ together with Lemma 5.4 as follows

$$\begin{aligned}(H_n)_*(M, 0) &= \min_{\sigma \in \mathbb{S}^{m-1}} H_n(M, \sigma) \\ &= \min_{\sigma \in \mathbb{S}^{m-1}} H_n(\Pi(\sigma)M\Pi(\sigma), 0) \\ &= \min_{\sigma \in \mathbb{S}^{m-1}} H_n((\Pi(\sigma)M\Pi(\sigma))_+, 0).\end{aligned}$$

□

In the case when $n \geq m-1$ we have that $k = m-1$, and in this case (5.11) says that when $p \neq 0$

$$H_n(M, p) = \operatorname{tr}((\Pi(p)M\Pi(p))_+).$$

For a given smooth function ϕ, the quantity $|\nabla\phi|^{-1} H_n(D^2\phi(x), \nabla\phi(x))$ is the sum of the non-negative principal curvatures of the level set of ϕ passing through x (this at least when $|\nabla\phi(x)| \neq 0$). If the sublevel sets of ϕ are convex, we have their mean curvature,

$$H_n(D^2\phi, \nabla\phi) = |\nabla\phi| \operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right).$$

5.2. **Flow by principal curvatures and flow by positive principal curvatures.** The last above point is important for us and is worth dwelling upon. Things are different if $u(\cdot, t)$ has sublevel sets which fail to be convex. To make this situation clear we consider another operator, defined for every n by

$$\overline{H}_n(M, p) = \sum_{j=0}^{k-1} \sigma_{m-j}(\Pi(p)M\Pi(p)),$$

with $k = \min\{n, m-1\}$ just as before (note the difference is we are no longer taking the positive part of $\Pi(p)M\Pi(p)$). We also consider the associated level set PDE,

$$\partial_t \overline{w} + \tfrac{1}{2}\overline{H}_n(D^2\overline{w}, \nabla\overline{w}) = 0. \tag{5.12}$$

This PDE is the level set formulation for flow by the sum of the largest k curvatures – in contrast to the flow by the sum of the largest k *non-negative* curvatures. We note that if $\overline{w}$ is a solution of (5.12) such that $\overline{w}(\cdot, t)$ has convex sublevel sets for every t, then $\overline{w}$ will also solve

$$\partial_t \overline{w} + \tfrac{1}{2} H_n(D^2\overline{w}, \nabla\overline{w}) = 0,$$

(the difference is we have the operator H_n instead of $\overline{H}_n$). Later at the end of Section 6 we review how when the terminal data is given by $\|\cdot\|_E$ the solution to (5.12) in fact has this property – in fact, for $n \geq m-1$ and smooth solutions this follows from a celebrated theorem by Huisken [69]. The corresponding extension of Huisken's theorem for viscosity solutions was obtained by Evans and Spruck [49], see also discussion at the end of Appendix B.

6. Proof of Theorem 2.1

Now we shall show the convergence of the rescaled value functions, $w_\delta : \mathbb{R}^m \times (\delta^2\mathbb{Z}_-) \to \mathbb{R}$, as $\delta \to 0$. As we will see, they converge to a continuous function $w(x,t)$ defined in $\mathbb{R}^m \times (-\infty, 0]$, and $w(x,t)$ is the unique viscosity solution to the terminal value problem

$$\begin{cases} \partial_t w + \frac{1}{2}H_n(D^2 w, \nabla w) & = 0 \text{ in } \mathbb{R}^m \times (-\infty, 0), \\ w(x,0) & = \|x\|_E \text{ in } \mathbb{R}^m. \end{cases} \tag{6.1}$$

In Appendix B we review the basic definitions and some essential results from the theory of viscosity solutions needed in this section.

The convergence happens in a locally uniform way (restricted to the domain of definition of w_δ), as detailed below. We record this as a theorem, the proof of which will be the main focus of this section.

Theorem 6.1. *Assume E has a C^2 boundary. As $\delta \to 0^+$, the functions $w_\delta(x,t)$ converge to a limit $w(x,t)$ in the sense that for any $R > 0$ we have*

$$\lim_{\delta \to 0^+} \max\{|w_\delta(x,t) - w(x,t)| \text{ for } |x| \leq R, |t| \leq R \text{ and } t \in \delta^2\mathbb{Z}_-\} = 0,$$

where $w : \mathbb{R}^m \times (-\infty, 0] \to \mathbb{R}$ is the unique viscosity solution of the terminal value problem (6.1).

6.1. **Setting up the proof of convergence.** The very first step towards showing the convergence is to bound the values $\{w_\delta(x,t)\}_\delta$ uniformly in δ at least in compact subsets of (x,t)'s.

Proposition 6.2. *Assume the convex body E has a C^2 boundary, there is a $C \equiv C(E,n)$ such that*

$$\|x\|_E \leq w_\delta(x,t) \leq \sqrt{\|x\|_E^2 - Ct},\ \forall\, x \in \mathbb{R}^m, t < 0.$$

Proof. This is simply a restating of Lemma 4.4 in terms of w_δ. Indeed, from the definition of w_δ and the pointwise bound from Lemma 4.4 we have

$$w_\delta(x,t) \leq \delta v(x/\delta, t/\delta^2) \leq \delta\sqrt{\|x/\delta\|_E^2 - C(t/\delta^2)} = \sqrt{\|x\|_E^2 - Ct}.$$

In light of this, note that

$$w_\delta(x,t) - \|x\|_E \leq \sqrt{\|x\|_E^2 - Ct} - \|x\|_E \leq \sqrt{C|t|}.$$

The last inequality follows from the inequality $\sqrt{a+b} - \sqrt{a} \leq \sqrt{b}$ valid whenever $a, b \geq 0$. The fact that $w_\delta(x,t) \geq \|x\|_E$ follows from the monotonicity of $v(y,s)$ with respect to s.

□

In light of the bound in Proposition 6.2 the convergence of the sequence $\{w_\delta\}_\delta$ would fail only if (informally speaking) the "limsup" and "liminf" of the sequence take different (but in any event finite) values. To investigate this, we must work with the half-relaxed limits of $\{w_\delta\}_\delta$, these are defined by

$$w^*(x,t) := \limsup_{\delta \to 0^+,\ x' \to x,\ t' \to t} w_\delta(x', t'),$$
$$w_*(x,t) := \liminf_{\delta \to 0^+,\ x' \to x,\ t' \to t} w_\delta(x', t'),$$

where the convergence is understood to be along arbitrary sequences $\{\delta_k\}_k, \{x_k'\}_k, \{t_k'\}_k$ such that

$$\delta_k > 0,\ t_k' \in \delta_k^2\mathbb{Z}_- \text{ for all } k.$$

The functions w^* and w_* are locally bounded functions in $\mathbb{R}^m \times (-\infty, 0]$. From their definition and from Proposition 6.2 it is clear that

$$\|x\|_E \leq w_*(x,t) \leq w^*(x,t) \leq \sqrt{\|x\|_E^2 - Ct}.$$

For a generic family of functions w_δ we cannot expect more than the above, and so $w_* \not\equiv w^*$ in general. Now we work w to rule out this last scenario, and here is where we use discrete equation solved by w_δ. Concretely, we will show w_* and w^* are respectively a viscosity supersolution and viscosity subsolution for the problem (6.1) (see Appendix B for a review of basic notions and results). From here, the comparison principle for viscosity sub/supersolutions says that

$$w^*(x,t) \leq w_*(x,t) \text{ for all } x, t,$$

from where $w_* = w^* = w$ will follow. The agreement between the half-relaxed limits guarantees the locally uniform convergence stated in Theorem 6.1, as we explain in the following remark.

Remark 6.3. *Consider $\delta_k \to 0$, and for each k take $(x_k, t_k) \in \overline{B}_R(0) \times [-R^2, 0] \cap \mathbb{R}^m \times \delta_k^2 \mathbb{Z}_-$ such that*

$$|w_{\delta_k}(x_k, t_k) - w(x_k, t_k)| = \max\{|w_{\delta_k}(x,t) - w(x,t)| \text{ for } |x| \le R, |t| \le R \text{ and } t \in \delta_k^2 \mathbb{Z}_-\}$$

Passing to a subsequence (δ'_k, x'_k, t'_k) we may assume without loss of generality $x'_k \to x_*$ and $t'_k \to t_*$ for some (x_*, t_*). Then, since w is continuous, we have that

$$\lim_k |w_{\delta'_k}(x'_k, t'_k) - w(x'_k, t'_k)| = |w(x_*, t_*) - w(x_*, t_*)| = 0.$$

6.2. **Establishing the half-relaxed limits are a sub/super solution.** We now proceed to show the subsolution and supersolution property for w^* and w_* respectively. First, we show both w^* and w_* have the right terminal values at $t = 0$ (in particular, the functions agree at $t = 0$).

Proposition 6.4. *As before, assume E has a C^2 boundary. The half-relaxed limits agree at time 0, namely*

$$w^*(x, 0) = w_*(x, 0) = \|x\|_E.$$

Proof. This is a consequence of the fact that $w_\delta(x,t)$ is monotone in t and the bound in Proposition 6.2. Indeed, on one hand $w_\delta(x, t_2) \ge w_\delta(x, t_1)$ for every $t_1, t_2 \in \delta^2 \mathbb{Z}_-$ with $t_2 \le t_1$, so that in particular

$$w_\delta(x,t) \ge w_\delta(x,0) = \|x\|_E \ \forall\, x \in \mathbb{R}^m, \ \forall\, t \in \delta^2 \mathbb{Z}_-.$$

On the other hand, thanks to the second inequality in Proposition 6.2 we have

$$w_\delta(x,t) - \|x\|_E \le \sqrt{C|t|} \ \forall\, x \in \mathbb{R}^m, \ \forall\, t \in \mathbb{Z}_-.$$

In other words, for all (x,t) in the domain of definition of w_δ we have

$$\|x\|_E \le w_\delta(x,t) \le \|x\|_E + \sqrt{C|t|}.$$

Then, passing to the limit along sequences $x' \to x$ and $t' \to 0$ we conclude that

$$\|x\|_E \le w_*(x,0) \le w^*(x,0) \le \|x\|_E,$$

and therefore $w^*(x,0) = w_*(x,0) = \|x\|_E$. ☐

Lemma 6.5. *Suppose $\phi(x,t)$ is a smooth function such that $w^* - \phi$ has a local maximum at (x_0, t_0), $t_0 < 0$, then*

$$\partial_t \phi(x_0,t_0) + \tfrac{1}{2}(H_n)^*(D^2\phi(x_0,t_0), \nabla\phi(x_0,t_0)) \ge 0.$$

If instead $w_* - \phi$ has a local minimum at (x_0, t_0) then

$$\partial_t \phi(x_0,t_0) + \tfrac{1}{2}(H_n)_*(D^2\phi(x_0,t_0), \nabla\phi(x_0,t_0)) \le 0.$$

Proof. As this is a standard argument in the theory of viscosity solutions we leave the proof of this lemma for later in Appendix A. ☐

We note that in the present setting the argument for Lemma 6.5 relies crucially on the relationship between $\mathcal{H}_{\delta,n}$, $H_{\delta,n}$, and H_n as established in Lemma 5.4 and Proposition 5.3, from these the rest of the proof follows common arguments in the literature.

Proof of Theorem 6.1. As explained earlier in this section, the theorem will be proved if we can show that $w_*(x,t) = w^*(x,t)$ for all x, t. Since E has a C^2 boundary we can apply Proposition 6.4 and Lemma 6.5 and conclude that (in the viscosity sense) we have

$$\partial_t w^* + \tfrac{1}{2} H_n(D^2 w^*, \nabla w^*) \ge 0 \text{ in } \mathbb{R}^m \times (-\infty, 0),$$
$$w^*(x,0) = \|x\|_E,$$

and

$$\partial_t w_* + \tfrac{1}{2} H_n(D^2 w_*, \nabla w_*) \le 0 \text{ in } \mathbb{R}^m \times (-\infty, 0),$$
$$w_*(x,0) = \|x\|_E,$$

Furthermore,

$$\|x\|_E \leq w_*(x,t) \leq w^*(x,t) \leq \sqrt{\|x\|_E^2 - Ct}.$$

We are then within the hypothesis of the comparison principle (Theorem B.2) and conclude that $w^* \leq w_*$, and so $w^* = w_*$. It follows w_* and w^* are the same function which is both a subsolution and a supersolution, and therefore is the unique viscosity solution to the terminal value problem. This finishes the proof. □

Remark 6.6. *Strictly speaking, the level sets of w are not evolving by the curvature flow due to the factor of $1/2$ in the PDE. If we set $\tilde{w}(x,t) := w(x,2t)$, then $\partial_t \tilde{w} = 2\partial_t w$ so that*

$$\partial_t \tilde{w} + H_n(D^2\tilde{w}, \nabla \tilde{w}) = 0,$$

and the level sets of the rescaled function $\tilde{w}$ are evolving via the respective curvature flow.

6.3. **Proof of the main theorem.** We recall the functional $\mathcal{K}_{n,T}(E)$, which we defined by

$$\mathcal{K}_{n,T}(E) = \max_{A_1} \min_{\bar{\varepsilon}_1} \ldots \max_{A_T} \min_{\bar{\varepsilon}_T} \|A_1\bar{\varepsilon}_1 + \ldots + A_T\bar{\varepsilon}_T\|_E.$$

At first sight it is not obvious the limit might exist. This convergence and identification of the limit will be established first when E has a smooth boundary (by means of Theorem 6.1). For a non smooth E we will use an approximation argument exploiting the monotonicity and homogeneity properties of the functionals $\mathcal{K}_{n,T}$ and τ_n. First, the monotonicity: if E, E' are centrally symmetric convex bodies and $E \subset E'$, then

$$\mathcal{K}_{n,T}(E') \leq \mathcal{K}_{n,T}(E),\ \tau_n(E) \leq \tau_n(E').$$

Second, the homogeneity: if E is a centrally symmetric convex body and $\lambda > 0$ we have

$$\mathcal{K}_{n,T}(\lambda E) = \lambda^{-1}\mathcal{K}_{n,T}(E),\ \tau_n(\lambda E) = \lambda^2 \tau_n(E).$$

These properties imply the following approximation lemma.

Lemma 6.7. *Given a centrally symmetric convex body E and $\epsilon \in (0,1)$ there is a centrally symmetric convex body E_ϵ with a smooth boundary such that for every n and $t \in \mathbb{N}$ we have*

$$\frac{1-\epsilon}{1+\epsilon}\mathcal{K}_{n,T}(E_\epsilon)\sqrt{\tau_n(E_\epsilon)} \leq \mathcal{K}_{n,T}(E)\sqrt{\tau_n(E)} \leq \frac{1+\epsilon}{1-\epsilon}\mathcal{K}_{n,T}(E_\epsilon)\sqrt{\tau_n(E_\epsilon)}.$$

Proof. For E as in the statement of the lemma and $\varepsilon > 0$ small, there is a centrally symmetric convex body E_ε with a smooth boundary such that

$$(1-\varepsilon)E_\varepsilon \subset E \subset (1+\varepsilon)E_\varepsilon.$$

From the reverse monotonicity of $\mathcal{K}_{n,T}$ we have

$$\mathcal{K}_{n,T}((1+\epsilon)E_\varepsilon) \leq \mathcal{K}_{n,T}(E) \leq \mathcal{K}_{n,T}((1-\varepsilon)E_\varepsilon),$$

and from the homogeneity of $\mathcal{K}_{n,T}$,

$$\frac{1}{1+\varepsilon}\mathcal{K}_{n,T}(E_\varepsilon) \leq \mathcal{K}_{n,T}(E) \leq \frac{1}{1-\varepsilon}\mathcal{K}_{n,T}(E_\varepsilon).$$

Likewise, for the τ_n functional, the inclusions imply that

$$\tau_n((1-\epsilon)E_\epsilon) \leq \tau_n(E) \leq \tau_n((1+\epsilon)E_\epsilon),$$

and then the homogeneity of τ_n yields

$$(1-\epsilon)\sqrt{\tau_n(E_\epsilon)} \leq \sqrt{\tau_n(E)} \leq (1+\epsilon)\sqrt{\tau_n(E_\epsilon)}.$$

Multiplying the inequalities for $\mathcal{K}_{n,T}$ and τ_n the lemma follows. □

Proof of Theorem 2.1. **Part I: Case when E has a smooth boundary.** In terms of the value function $v = v_E$ defined by (4.1), we have the identity

$$\frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}} = \frac{v(0,-T)}{\sqrt{T}}.$$

Now given that $w_\delta(x,t) = \delta v(x/\delta, t/\delta^2)$ by definition we also have

$$\frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}} = \frac{\delta^{-1} w_\delta(0,-\delta^2 T)}{\sqrt{T}}.$$

Choosing $\delta = 1/\sqrt{T}$, this results in

$$\frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}} = w_{1/\sqrt{T}}(0,-1).$$

Given the smoothness assumption on E we can apply Theorem 6.1 and conclude $w_\delta \to w$ locally uniformly, and in particular $w_{1/\sqrt{T}}(0,-1) \to w(0,-1)$ as $T \to \infty$, where $w(x,t)$ is the unique solution to (6.1). This establishes the existence of the limit

$$\lim_{T\to\infty} \frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}},$$

which agrees with $w(0,-1)$. Now we consider the function $\overline{w}$ solving the terminal value problem

$$\partial_t \overline{w} + \tfrac{1}{2}\overline{H}_n(D^2\overline{w}, \nabla\overline{w}) = 0 \text{ in } \mathbb{R}^m \times \mathbb{R}_-$$
$$\overline{w}(x,0) = \|x\|_E,$$

that is, $\overline{w}$ is the level set function describing flow by the sum of the largest $\min\{n, m-1\}$ curvatures. Then, according to Theorem B.4 (see discussion at the end of Appendix B) the unique viscosity solution $\overline{w}$ has the property that $\overline{w}(\cdot,t)$ has convex sublevel sets for every t. In light of this (recall the discussion in Section 5.2) it can be shown that $\overline{w}$ is also a viscosity solution of

$$\partial_t \overline{w} + \tfrac{1}{2} H_n(D^2\overline{w}, \nabla\overline{w}) = 0.$$

Then, by the comparison principle for viscosity solutions (see Theorem B.2) we must have $w(x,t) = \overline{w}(x,t)$ for all x, t. In particular (recall Remark 6.6) the level sets of $w(\cdot, 2t)$ are evolving according to the sum of the largest $\min\{n, m-1\}$ principal curvatures.

To conclude this first part of the proof it remains to explain why is it that

$$w(0,-1) = \frac{1}{\sqrt{2\tau_n(E)}}. \tag{6.2}$$

For this, we find it convenient to argue in terms of the rescaled function $\tilde{w}(x,t) = w(x,2t)$, whose level sets are evolving by the curvature flow (again, as noted in Remark 6.6). In particular, we see for $t > 0$ the value $\tilde{w}(0,-t)$ is equal that of the level set of $\tilde{w}(\cdot,0) = \|\cdot\|_E$ with extinction time equal to t. Since $E = \{\tilde{w}(\cdot,0) \le 1\}$, it follows the level set vanishing at time $\tau_n(E)$ is the one corresponding to $\{\tilde{w}(\cdot,0) = 1\}$, that is ∂E. It follows that $\tilde{w}(0,-\tau_n(E)) = 1$. Now, the fact that $w(\cdot,0)$ is homogeneous of degree 1 means $w(x,t)$ will be self similar, and one has that (see Lemma B.3)

$$w(x,t) = \sqrt{-t}\, w(x/\sqrt{-t}, -1).$$

Therefore,

$$1 = \tilde{w}(0,-\tau_n(E)) = w(0,-2\tau_n(E)) = \sqrt{2\tau_n(E)}\, w(0,-1),$$

and the identity (6.2) follows. This proves that

$$\lim_{T\to\infty} \frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}} = \frac{1}{\sqrt{2\tau_n(E)}},$$

in the case when ∂E is smooth.

Part II: Approximation of general E by smooth convex bodies. For the general case fix E and $\epsilon \in (0,1)$. Then by Lemma 6.7 there is a smooth, centrally symmetric convex body E_ϵ such that

$$\frac{1-\epsilon}{1+\epsilon}\mathcal{K}_{n,T}(E_\epsilon)\sqrt{\tau_n(E_\epsilon)} \le \mathcal{K}_{n,T}(E)\sqrt{\tau_n(E)} \le \frac{1+\epsilon}{1-\epsilon}\mathcal{K}_{n,T}(E_\epsilon)\sqrt{\tau_n(E_\epsilon)}.$$

The convex body E_ε has a smooth boundary and thus according to Part I,

$$\lim_{T\to\infty}\left\{\frac{\mathcal{K}_{n,T}(E_\epsilon)}{\sqrt{T}}\sqrt{2\tau_n(E_\epsilon)}\right\} = 1.$$

Combining this limit with the inequalities above we conclude that

$$\frac{1-\epsilon}{1+\epsilon}\frac{1}{\sqrt{2\tau_n(E)}} \leq \liminf_{T\to\infty}\frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}} \leq \limsup_{T\to\infty}\frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}} \leq \frac{1+\epsilon}{1-\epsilon}\frac{1}{\sqrt{2\tau_n(E)}}.$$

Since this holds for all $\epsilon \in (0,1)$ we may take $\epsilon \to 0$ and conclude that

$$\lim_{T\to\infty}\frac{\mathcal{K}_{n,T}(E)}{\sqrt{T}} = \frac{1}{\sqrt{2\tau_n(E)}},$$

and the theorem is proved. □

7. Proof of Theorem 2.2

We now establish lower and upper bounds on the extinction time for E under mean curvature flow, in particular, $n \geq m-1$ throughout this section. To obtain the bounds, we will construct adequate subsolutions and supersolutions for the PDE solved by w, and then apply the comparison principle in the viscosity framework [49, Theorem 3.2]. Then, the bound for the extinction times follows from (6.2). As noted previously, the definitions of viscosity solutions and the comparison principle are reviewed in Appendix B.

Throughout this section we will study the specific terminal value problem

$$\begin{cases} \partial_t w + \frac{1}{2}|\nabla w|\,\mathrm{div}\left(\frac{\nabla w}{|\nabla w|}\right) & = 0 \text{ in } \mathbb{R}^m \times (-\infty, 0) \\ w(x,0) & = \|x\|_E \text{ in } \mathbb{R}^m \end{cases} \tag{7.1}$$

The upper bound will be the simpler of the two estimates, so we discuss the lower bound first.

7.1. **The lower bound.** For the lower bound we will show that a solution to the heat equation with a properly chosen diffusivity and terminal data will produce a subsolution to (7.1). Therefore, we define $\psi(x,t)$ as the solution to the terminal value problem,

$$\begin{cases} \partial_t\psi + \frac{1}{4}\Delta\psi & = 0 \text{ in } \mathbb{R}^m \times (-\infty, 0), \\ \psi(x,0) & = C(E)\max_i x_i, \end{cases} \tag{7.2}$$

where $C(E)$ is the constant defined by

$$C(E) := 1/\max_i \|e_i\|_E^*.$$

Remark 7.1. *The idea behind the definition of $C(E)$ comes from the case where $E = [-1,1]^m$. Then $\|x\|_E = \|x\|_\infty$ and we have that $\|x\|_\infty \geq \max x_i$, so $C_E = 1$ in this case.*

The initial data $g(x) := C(E)\max_i x_i$ has a few properties that will be useful going forward. First, $g(x)$ is convex and monotone increasing in each variable x_i, this means that if p is any element of its sub-differential, then the components of p are all non-negative.

Next, $g(x)$ is a *submodular* function, which means that if $e_1, \ldots, e_m$ denotes the canonical basis, then for any $h > 0$ and indices $i \neq j$ we have that

$$g(x + he_i + he_j) + g(x) - g(x+he_i) - g(x+he_j) \leq 0 \;\forall\, x \in \mathbb{R}^m.$$

Observe that a twice differentiable function is submodular if and only if the mixed second derivatives are all non-positive. We also observe $g(x)$ is linear along the direction of the diagonal. In other words, if $\mathbf{1} := (1,\ldots,1) \in \mathbb{R}^m$, then for every $x \in \mathbb{R}^m$ and $s \in \mathbb{R}$ we have

$$g(x + s\mathbf{1}) = g(x) + C(E)s.$$

These are all properties preserved by translations and convex combinations, and thus they are preserved by convolution against a non-negative smooth kernel. Since $\psi(x,t)$ is obtained by convolution of $g(x)$ against the heat kernel, the above applies to $\psi(x,t)$ for $t < 0$. These properties of $\psi(x,t)$ for $t < 0$ say that $-D^2\psi(x,t)$ is what is known as a Laplacian matrix.

Definition 7.2. *We will say a $L \in \mathbb{R}^{m\times m}$ is a Laplacian matrix if it is a symmetric matrix such that*[7]

$$L_{ij} \geq 0 \text{ if } i \neq j \text{ and } \sum_{j=1}^{m} L_{ij} = 0 \,\forall\, i \in [m].$$

This property yields a useful pointwise bound for $(D^2\psi\nabla\psi, \nabla\psi)$.

Lemma 7.3. *If $L \in \mathbb{R}^{m\times m}$ is a Laplacian matrix and p is a vector in $\mathbb{R}^m$ with $p_i \geq 0$ for each $i \in [m]$, then*

$$-(Lp, p) \leq \max_i\{-L_{ii}\}|p|^2.$$

Proof. Since $(Lp, p) = \sum_{ij} L_{ij}p_ip_j$ we use the fact that $p_i \geq 0$ and that L is a Laplacian matrix to obtain

$$-(Lp, p) = -\sum_i L_{ii}p_i^2 - \sum_i \sum_{j\neq i} L_{ij}p_ip_j \leq \max_i\{-L_{ii}\}|p|^2.$$

□

Lemma 7.4. *If ψ is the solution to (7.2) and w the solution to (7.1), then*

$$\psi(x, t) \leq w(x, t) \,\forall\, x \in \mathbb{R}^m,\ t < 0.$$

Proof. As observed above $-D^2\psi(x, t)$ is a Laplacian matrix for every $x \in \mathbb{R}^m$ and $t < 0$. Likewise, $\psi(x, t)$ is monotone in each variable x_i, and so $\partial_k\psi \geq 0$ for every k and $\psi(x, t)$ is linear in the direction of the vector of all ones, and therefore $\sum_k \partial_k\psi = C(E)$. Therefore, $\nabla\psi \neq 0$ everywhere for $t < 0$. In this case, we may apply Lemma 7.3 with $L = -D^2\psi$ and $p = \nabla\psi$ to obtain

$$\frac{1}{|\nabla\psi|^2}(D^2\psi\nabla\psi, \nabla\psi) \leq \max_k\{\partial_{kk}\psi\}, \quad \text{for } t < 0.$$

In particular, for all x and for all $t < 0$ we have

$$\Delta\psi - \frac{1}{|\nabla\psi|^2}(D^2\psi\nabla\psi, \nabla\psi) \geq \Delta\psi - \max_k\{\partial_{kk}\psi\}. \tag{7.3}$$

On the other hand, using again that $-D^2\psi$ is a Laplacian matrix, and therefore $\partial_{\ell j}\psi \leq 0$ for $\ell \neq j$, we have

$$\partial_{kk}\psi = \Delta\psi - \partial_{kk}\psi + \sum_{\ell\neq k}\sum_{j\neq k,\ell} \partial_{\ell j}\psi \leq \Delta\psi - \partial_{kk}\psi.$$

In particular, since $\partial_{kk}\psi \geq 0$, the preceding inequality guarantees

$$\Delta\psi = \Delta\psi - \partial_{kk}\psi + \partial_{kk}\psi \leq 2(\Delta\psi - \partial_{kk}\psi).$$

As this holds for every k, we have

$$\frac{1}{2}\Delta\psi \leq \Delta\psi - \max_k \partial_{kk}\psi \tag{7.4}$$

Now, we put the above inequalities together: the standard identity for the MCF operator and the inequality (7.3) provide that

$$\partial_t\psi + \tfrac{1}{2}|\nabla\psi| \operatorname{div}\left(\frac{\nabla\psi}{|\nabla\psi|}\right) = \partial_t\psi + \tfrac{1}{2}\Big(\Delta\psi - \frac{1}{|\nabla\psi|^2}(D^2\psi\nabla\psi, \nabla\psi)\Big) \geq \partial_t\psi + \tfrac{1}{2}\left(\Delta\psi - \max_k \partial_{kk}\psi\right),$$

and then the inequality (7.4) yields

$$\partial_t\psi + \tfrac{1}{2}|\nabla\psi| \operatorname{div}\left(\frac{\nabla\psi}{|\nabla\psi|}\right) \geq \partial_t\psi + \tfrac{1}{4}\Delta\psi = 0.$$

Finally, we obtain

$$C(E)\max_i x_i = \psi(x, 0) \leq \|x\|_E$$

from the following inequality based on definition of the dual norm (see Section 1.2):

$$\max_i x_i = \|x\|_E \max_i(e_i, x/\|x\|_E) \leq \|x\|_E \max_i \|e_i\|_E^*.$$

We conclude that ψ is a subsolution of (7.1), and thus $\psi \leq w$ by the comparison principle (Theorem B.2).

[7] As a reminder in our sign convention Laplacian is negative definite matrix. See footnote 6 in Section 3 regarding this convention.

□

Theorem 7.5. *The function $w(x,t)$ solving the problem (7.1) satisfies the following bound for every $T>0$*

$$C(E)\,\mathbb{E}[\max_i G_i]\sqrt{T/2}\,\leq w(0,-T).$$

where G denotes an m-dimensional standard Gaussian random vector with mean 0 and identity covariance.

Proof. From Lemma 7.4, we have that

$$\psi(0,-T)\leq w(0,-T).$$

and since ψ is a solution of the heat equation with diffusion constant $1/4$, we have

$$\psi(0,-T)=C(E)\mathbb{E}[\max_i G_i]\,\sqrt{T/2}.$$

□

7.2. **The upper bound.** As we mentioned earlier in the section, the corresponding upper bound for w is more straightforward than the lower bound. In this case, we simply make use of the fact that $\overline{\psi}(x,t)$ solving the backwards heat equation with terminal data $\|x\|_E$ serves as a supersolution to (7.1). The value of $\overline{\psi}(x,t)$ is then the expected norm of a Gaussian vector.

We start with the supersolution property for $\overline{\psi}$. For $t>0$ and $x\in\mathbb{R}^m$ we define $\overline{\psi}$ by

$$\overline{\psi}(x,-t):=\mathbb{E}[\|x+X_t\|_E],$$

where X_t is the standard m-dimensional Brownian motion starting from $X_0=0$. In particular, $\overline{\psi}$ solves

$$\begin{aligned}\partial_t\overline{\psi}+\tfrac{1}{2}\Delta\overline{\psi}&=0 \text{ in } \mathbb{R}^m\times(-\infty,0),\\ \overline{\psi}(x,0)&=\|x\|_E.\end{aligned}$$

Proposition 7.6. *If $w(x,t)$ is the solution of (7.1) and $\overline{\psi}$ is as defined above, then for every $t>0$ we have*

$$w(x,-t)\leq\overline{\psi}(x,-t).$$

Proof. As a solution of the heat equation $\overline{\psi}(\cdot,-t)$ is given by the convolution of $\|x\|_E$ with the heat kernel at time t. In particular, $\overline{\psi}(\cdot,-t)$ is twice differentiable, and as $\|x\|_E$ is convex, so will $\overline{\psi}(\cdot,-t)$.

Now, take a point (x_0,t_0) with $t_0<0$ such that $\nabla\overline{\psi}(x_0,t_0)\neq 0$. Then at (x_0,t_0) we have

$$\Delta\overline{\psi}=|\nabla\overline{\psi}|\operatorname{div}\left(\frac{\nabla\overline{\psi}}{|\nabla\overline{\psi}|}\right)+\frac{1}{|\nabla\overline{\psi}|^2}((D^2\overline{\psi})\nabla\overline{\psi},\nabla\overline{\psi})\geq|\nabla\overline{\psi}|\operatorname{div}\left(\frac{\nabla\overline{\psi}}{|\nabla\overline{\psi}|}\right),$$

thanks to $((D^2\overline{\psi})\nabla\overline{\psi},\nabla\overline{\psi})\geq 0$ everywhere. This shows that

$$\partial_t\overline{\psi}+\tfrac{1}{2}|\nabla\overline{\psi}|\operatorname{div}\left(\frac{\nabla\overline{\psi}}{|\nabla\overline{\psi}|}\right)\leq\partial_t\overline{\psi}+\tfrac{1}{2}\Delta\overline{\psi}=0,\ \text{ in } \{\nabla\overline{\psi}\neq 0\}.$$

If $\nabla\overline{\psi}(x_0,t_0)=0$, then at (x_0,t_0) we have (recall $(H_n)_*$ from Lemma 5.4 and Remark 5.5)

$$(H_n)_*(D^2\overline{\psi},0)=\min_{\sigma\in\mathbb{S}^{m-1}}H_n(D^2\overline{\psi},\sigma),$$

Since $n\geq m-1$ and for any $\sigma\in\mathbb{S}^{m-1}$ and M the matrix $\Pi(\sigma)M\Pi(\sigma)$ has at least zero as an eigenvalue, we have $H_n(M,\sigma)=\operatorname{tr}((\Pi(\sigma)M\Pi(\sigma))_+)$, and since $D^2\psi\geq 0$ we have $\Pi(\sigma)D^2\overline{\psi}\Pi(\sigma)\geq 0$. It follows that for

$$H_n(D^2\overline{\psi},\sigma)=\operatorname{tr}(\Pi(\sigma)D^2\overline{\psi}\Pi(\sigma))=\Delta\overline{\psi}-((D^2\overline{\psi})\sigma,\sigma)\leq\Delta\overline{\psi},$$

for any $\sigma\in\mathbb{S}^{m-1}$. In particular, $(H_n)_*(D^2\overline{\psi},0)\leq\Delta\overline{\psi}$. Therefore, at (x_0,t_0) we have also

$$\partial_t\overline{\psi}+\tfrac{1}{2}(H_n)_*(D^2\overline{\psi},0)\leq\partial_t\overline{\psi}+\tfrac{1}{2}\Delta\overline{\psi}=0.$$

We conclude $\overline{\psi}$ is a supersolution for the problem solved by w (we already have $\overline{\psi}=w$ at $t=0$ by definition). Therefore, we may apply the comparison principle (Theorem B.2) and conclude that $w(x,t)\leq\overline{\psi}(x,t)$ for all $x\in\mathbb{R}^m$ and $t<0$.

□

Thus, Proposition 7.6 provides the desired upper bound, the expected norm of an m-dimensional normal random variable is well understood:

$$w(0,-T) \le \overline{\psi}(0,-T) = \mathbb{E}\|X_t\|_E = \sqrt{T}\,\mathbb{E}\|G\|_E$$

where G denotes an m-dimensional standard Gaussian random vector with mean 0 and identity covariance.

Acknowledgements. We honor the memory of Robert Kohn whose constant encouragement and generous sharing of ideas made this project possible. V.A.K. expresses his profound appreciation to Afonso Bandeira for the introduction to the Komlós conjecture, which was the initial spark for this project. V.A.K. is immensely grateful to his postdoctoral mentor, Maria Han Veiga, for her guidance and collaboration throughout his time at The Ohio State University. N.G. is indebted to Deane Yang and Shay Sadovsky for many helpful and interesting discussions. N.G. would like to acknowledge the generous support of the National Science Foundation through grant DMS-214423.

References

[1] Elad Aigner-Horev, Dan Hefetz, and Michael Trushkin. Smoothed analysis of the Komlós conjecture: Rademacher noise. *The Electronic Journal of Combinatorics*, 32(1):P1.52, Mar. 2025.

[2] Alex Albors, Hisham Bhatti, Lukshya Ganjoo, Raymond Guo, Dmitriy Kunisky, Rohan Mukherjee, Alicia Stepin, and Tony Zeng. On the structure of bad science matrices, 2025.

[3] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. John Wiley & Sons, Inc., 2000.

[4] Dylan J. Altschuler and Jonathan Niles-Weed. The discrepancy of random rectangular matrices. *Random Structures & Algorithms*, 60(4):551–593, 2022.

[5] Dylan J. Altschuler and Konstantin Tikhomirov. A threshold for online balancing of sparse i.i.d. vectors, 2025.

[6] Ryan Alweiss, Yang P. Liu, and Mehtaab Sawhney. Discrepancy minimization via a self-balancing walk. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 14–20, New York, NY, USA, 2021. Association for Computing Machinery.

[7] Ben Andrews. Contraction of convex hypersurfaces in Euclidean space. *Calculus of Variations and Partial Differential Equations*, 2(2):151–171, 1994.

[8] Benjamin Aubin, Will Perkins, and Lenka Zdeborová. Storage capacity in symmetric binary perceptrons. *Journal of Physics A: Mathematical and Theoretical*, 52(29):294003, June 2019.

[9] Wojciech Banaszczyk. A Beck—Fiala-type theorem for Euclidean norms. *European Journal of Combinatorics*, 11(6):497–500, 1990.

[10] Wojciech Banaszczyk. Balancing vectors and convex bodies. *Studia Mathematica*, 106(1):93–100, 1993.

[11] Wojciech Banaszczyk. Balancing vectors and Gaussian measures of n-dimensional convex bodies. *Random Structures & Algorithms*, 12(4):351–360, 1998.

[12] Afonso S. Bandeira and Helmut Bölcskei. Matrix discrepancy for representations of finite groups, 2026.

[13] Afonso S. Bandeira, Dmitriy Kunisky, Dustin G. Mixon, and Xinmeng Zeng. On the concentration of Gaussian Cayley matrices. *Applied and Computational Harmonic Analysis*, 73:101694, 2024.

[14] Afonso S Bandeira, Antoine Maillard, and Nikita Zhivotovskiy. A remark on Kashin's discrepancy argument and partial coloring in the Komlós conjecture. *Portugaliae Mathematica*, 79(3):311–316, 2022.

[15] Nikhil Bansal. Constructive algorithms for discrepancy minimization. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 3–10, 2010.

[16] Nikhil Bansal. Discrepancy theory and related algorithms. In *Proc. Int. Cong. Math*, volume 7, pages 5178–5210. International Mathematical Union, 2022.

[17] Nikhil Bansal, Daniel Dadush, and Shashwat Garg. An algorithm for Komlós conjecture matching Banaszczyk's bound. *SIAM Journal on Computing*, 48(2):534–553, 2019.

[18] Nikhil Bansal, Daniel Dadush, Shashwat Garg, and Shachar Lovett. The Gram–Schmidt walk: A cure for the Banaszczyk blues. *Theory of Computing*, 15(21):1–27, 2019.

[19] Nikhil Bansal and Haotian Jiang. Decoupling via affine spectral-independence: Beck-Fiala and Komlós bounds beyond Banaszczyk. In *Proceedings of the 58th Annual ACM SIGACT Symposium on Theory of Computing, to appear*, STOC 2026, New York, NY, USA, 2026. Association for Computing Machinery.

[20] Nikhil Bansal, Haotian Jiang, Raghu Meka, Sahil Singla, and Makrand Sinha. Online discrepancy minimization for stochastic arrivals. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2842–2861.

[21] Nikhil Bansal, Haotian Jiang, Raghu Meka, Sahil Singla, and Makrand Sinha. Prefix Discrepancy, Smoothed Analysis, and Combinatorial Vector Balancing. In Mark Braverman, editor, *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, volume 215 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 13:1–13:22, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[22] Nikhil Bansal, Haotian Jiang, Raghu Meka, Sahil Singla, and Makrand Sinha. Smoothed Analysis of the Komlós Conjecture. In Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff, editors, *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, volume 229 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 14:1–14:12, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[23] Nikhil Bansal, Haotian Jiang, Sahil Singla, and Makrand Sinha. Online vector balancing and geometric discrepancy. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2020, page 1139–1152, New York, NY, USA, 2020. Association for Computing Machinery.

[24] Nikhil Bansal, Aditi Laddha, and Santosh Vempala. A Unified Approach to Discrepancy Minimization. In Amit Chakrabarti and Chaitanya Swamy, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2022)*, volume 245 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1:1–1:22, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[25] Nikhil Bansal and Joel H. Spencer. On-line balancing of random inputs. *Random Structures & Algorithms*, 57(4):879–891, 2020.

[26] Martino Bardi, Italo Capuzzo Dolcetta, et al. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*, volume 12. Springer, 1997.

[27] G. Barles and P. E. Souganidis. Convergence of approximation schemes for fully nonlinear second order equations. *Asymptotic Anal.*, 4(3):271–283, 1991.

[28] József Beck and Tibor Fiala. "Integer-making" theorems. *Discrete Applied Mathematics*, 3(1):1–8, 1981.

[29] Witold Bednorz and Piotr Godlewski. Some remarks on the Gram-Schmidt walk algorithm and consequences for the Komlós conjecture. In Radosław Adamczak, Nathael Gozlan, Karim Lounici, Mokshay Madiman, Florence Merlevède, and Elisabeth Werner, editors, *High Dimensional Probability X*, pages 37–66, Cham, 2026. Springer Nature Switzerland.

[30] Richard Bellman. *Dynamic programming*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 2010. Reprint of the 1957 edition, With a new introduction by Stuart Dreyfus.

[31] Kenneth A. Brakke. *The motion of a surface by its mean curvature*. Mathematical Notes. Princeton Univ. Press, Princeton, NJ, 1978.

[32] Rainer Buckdahn, Pierre Cardaliaguet, and Marc Quincampoix. A representation formula for the mean curvature motion. *SIAM Journal on Mathematical Analysis*, 33(4):827–846, 2001.

[33] Jeff Calder. Lecture notes on viscosity solutions, 2024. Available at https://www-users.cse.umn.edu/~jwcalder/viscosity_solutions.pdf.

[34] Karthekeyan Chandrasekaran and Santosh S. Vempala. Integer feasibility of random polytopes: random integer programs. In *Proceedings of the 5th Conference on Innovations in Theoretical Computer Science*, ITCS '14, page 449–458, New York, NY, USA, 2014. Association for Computing Machinery.

[35] Moses Charikar, Alantha Newman, and Aleksandar Nikolov. Tight hardness results for minimizing discrepancy. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, page 1607–1614, USA, 2011. Society for Industrial and Applied Mathematics.

[36] Fernando Charro, Jesus García Azorero, and Julio D Rossi. A mixed problem for the infinity laplacian via tug-of-war games. *Calculus of Variations and Partial Differential Equations*, 34(3):307–320, 2009.

[37] Giorgos Chasapis and Nikos Skarmogiannis. A note on norms of signed sums of vectors. *Advances in Geometry*, 21(1):5–14, 2021.

[38] Bernard Chazelle. *The Discrepancy Method: Randomness and Complexity*. Cambridge University Press, 2000.

[39] J. Cheeger. A lower bound for the smallest eigenvalue of the Laplacian. Probl. Analysis, Sympos. in Honor of Salomon Bochner, Princeton Univ. 1969, 195-199 (1970)., 1970.

[40] William Chen, Anand Srivastav, and Giancarlo Travaglini, editors. *A Panorama of Discrepancy Theory*. Springer Cham, 2014.

[41] Yun Gang Chen, Yoshikazu Giga, and Shun'ichi Goto. Uniqueness and existence of viscosity solutions of generalized mean curvature flow equations. *Proc. Japan Acad. Ser. A Math. Sci.*, 65(7):207–210, 1989.

[42] Sinho Chewi, Patrik Gerber, Philippe Rigollet, and Paxton Turner. Gaussian discrepancy: A probabilistic relaxation of vector balancing. *Discrete Applied Mathematics*, 322:123–141, 2022.

[43] Tobias Colding, William Minicozzi, Erik Pedersen, et al. Mean curvature flow. *Bulletin of the American Mathematical Society*, 52(2):297–333, 2015.

[44] Kevin P. Costello. Balancing Gaussian vectors. *Israel Journal of Mathematics*, 172(1):145–156, 2009.

[45] Philippe Courrege. Sur la forme intégro-différentielle des opérateurs de c_k^∞ dans c satisfaisant au principe du maximum. *Séminaire de théorie du potentiel*, 10(1):1–38, 1965.

[46] Michael G Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User's guide to viscosity solutions of second order partial differential equations. *Bulletin of the American mathematical society*, 27(1):1–67, 1992.

[47] Anirban DasGupta, S. N. Lahiri, and Jordan Stoyanov. Sharp fixed n bounds and asymptotic expansions for the mean and the median of a Gaussian sample maximum, and applications to the Donoho–Jin model. *Statistical Methodology*, 20:40–62, 2014.

[48] Ronen Eldan and Mohit Singh. Efficient algorithms for discrepancy minimization in convex sets. *Random Structures & Algorithms*, 53(2):289–307, 2018.

[49] Lawrence C Evans and Joel Spruck. Motion of level sets by mean curvature. I. In *Fundamental Contributions to the Continuum Theory of Evolving Phase Interfaces in Solids: A Collection of Reprints of 14 Seminal Papers*, pages 328–374. Springer, 1991.

[50] Esther Ezra and Shachar Lovett. On the Beck-Fiala conjecture for random set systems. *Random Structures & Algorithms*, 54(4):665–675, 2019.

[51] Christian Fiedler, Joe Jackson, Daniel Lacker, and Jonathan Niles-Weed. The mean-field limit of online stochastic vector balancing, 2026.

[52] Wendell H Fleming and H Mete Soner. *Controlled Markov processes and viscosity solutions*. Springer, 2006.

[53] Cole Franks and Michael Saks. On the discrepancy of random matrices with many columns. *Random Structures & Algorithms*, 57(1):64–96, 2020.
[54] J. Galambos. *Asymptotic Theory of Extreme Order Statistics*. Wiley, New York, 1987.
[55] Apostolos Giannopoulos. On some vector balancing problems. *Studia Mathematica*, 122(3):225–234, 1997.
[56] Y. Giga, S. Goto, H. Ishii, and M.-H. Sato. Comparison principle and convexity preserving properties for singular degenerate parabolic equations on unbounded domains. *Indiana Univ. Math. J.*, 40(2):443–470, 1991.
[57] Yoshikazu Giga. *Surface evolution equations: A level set approach*. Springer, 2006.
[58] Yoshikazu Giga and Qing Liu. *Surface Evolution Equations: A Level Set Approach*. Monographs in Mathematics. Birkhäuser Cham, 2 edition, 2026.
[59] Yoshikazu Giga and Kazuyuki Yama-uchi. On a lower bound for the extinction time of surfaces moved by mean curvature. *Calculus of Variations and Partial Differential Equations*, 1(4):417–428, 1993.
[60] Efim Davydovich Gluskin. Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces. *Mathematics of the USSR-Sbornik*, 64(1):85–96, 1989.
[61] Irene Gonzalvez, Alfredo Miranda, Julio D. Rossi, and Jorge Ruiz-Cases. A two-player zero-sum probabilistic game that approximates the mean curvature flow. *Communications in Mathematics*, Volume 34 (2026), Issue 2 (Special issue: Latin American mathematics), Jul 2025.
[62] Nestor Guillen and Russell W. Schwab. Min-max formulas for nonlocal elliptic operators. *Calc. Var. Partial Differential Equations*, 58(6):Paper No. 209, 79, 2019.
[63] Nestor Guillen and Russell W. Schwab. Min-max formulas for nonlocal elliptic operators on Euclidean space. *Nonlinear Anal.*, 193:111468, 51, 2020.
[64] C. Sinan Güntürk. Mathematics of analog-to-digital conversion. *Communications on Pure and Applied Mathematics*, 65(12):1671–1696, 2012.
[65] Nika Haghtalab, Tim Roughgarden, and Abhishek Shetty. Smoothed analysis with adaptive adversaries. *J. ACM*, 71(3), June 2024.
[66] D Hajela. On a conjecture of Komlos about signed sums of vectors inside the sphere. *European Journal of Combinatorics*, 9(1):33–37, 1988.
[67] Christopher Harshaw, Fredrik Sävje, Daniel A. Spielman, and Peng Zhang. Balancing covariates in randomized experiments with the Gram–Schmidt walk design. *Journal of the American Statistical Association*, 119(548):2934–2946, 2024.
[68] Rebecca Hoberg and Thomas Rothvoss. A Fourier-analytic approach for the discrepancy of random set systems. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '19, page 2547–2556, USA, 2019. Society for Industrial and Applied Mathematics.
[69] Gerhard Huisken. Flow by mean curvature of convex surfaces into spheres. *Journal of Differential Geometry*, 20(1):237–266, 1984.
[70] Cyril Imbert and Sylvia Serfaty. Repeated games for non-linear parabolic integro-differential equations and integral curvature flows. *Discrete and Continuous Dynamical Systems*, 29(4):1517–1552, 2011.
[71] Haotian Jiang, Janardhan Kulkarni, and Sahil Singla. Online geometric discrepancy for stochastic arrivals with applications to envy minimization, 2019.
[72] Gautam Kamath. Bounds on the expectation of the maximum of samples from a Gaussian, 1998.
[73] B. S. Kashin. On an isometric operator in $L^2(0,1)$. *C. R. Acad. Bulg. Sci.*, 38:1613–1615, 1985.
[74] Vladimir A. Kobzar. The symmetric two-armed bandit: the general case, 2026.
[75] Vladimir A. Kobzar and Robert V. Kohn. A PDE-based analysis of the symmetric two-armed Bernoulli bandit, 2022.
[76] Vladimir A. Kobzar, Robert V. Kohn, and Zhilei Wang. New potential-based bounds for prediction with expert advice. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of the 33rd Annual Conference on Learning Theory (COLT)*, volume 125 of *Proceedings of Machine Learning Research*, pages 2370–2405. PMLR, 09–12 Jul 2020.
[77] Vladimir A. Kobzar, Robert V. Kohn, and Zhilei Wang. New potential-based bounds for the geometric-stopping version of prediction with expert advice. In Jianfeng Lu and Rachel Ward, editors, *Proceedings of the 1st Annual Conference on Mathematical and Scientific Machine Learning*, volume 107 of *Proceedings of Machine Learning Research*, pages 537–554, Princeton University, Princeton, NJ, USA, 20–24 Jul 2020. PMLR.
[78] Robert Kohn and Sylvia Serfaty. A deterministic-control-based approach to motion by curvature. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(3):344–407, 2006.
[79] Robert V. Kohn and Sylvia Serfaty. Second-order PDE's and deterministic games. In *ICIAM 07 : 6th International Conference on Industrial and Applied Mathematics*, Zurich, Switzerland, July 2007.
[80] Robert V. Kohn and Sylvia Serfaty. A deterministic-control-based approach to fully nonlinear parabolic and elliptic equations. *Communications on Pure and Applied Mathematics*, 63:1298–1350, 2010.
[81] Janardhan Kulkarni, Victor Reis, and Thomas Rothvoss. Optimal online discrepancy minimization. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 1832–1840, New York, NY, USA, 2024. Association for Computing Machinery.
[82] Dmitriy Kunisky. The discrepancy of unsatisfiable matrices and a lower bound for the Komlós conjecture constant. *SIAM Journal on Discrete Mathematics*, 37(2):586–603, 2023.
[83] Dmitriy Kunisky, Timm Oertel, Nicola Wengiel, and Peiyuan Zhang. Asymptotic bounds and online algorithms for average-case matrix discrepancy, 2025.
[84] Martin Larsson and Johannes Ruf. Minimum curvature flow and martingale exit times. *Electronic Journal of Probability*, 29:1–32, 2024.

[85] Avi Levy, Harishchandra Ramadas, and Thomas Rothvoss. Deterministic discrepancy minimization via the multiplicative weight update method. In Friedrich Eisenbrand and Jochen Koenemann, editors, *Integer Programming and Combinatorial Optimization*, pages 380–391, Cham, 2017. Springer International Publishing.
[86] Qing Liu, Armin Schikorra, and Xiaodan Zhou. A game-theoretic proof of convexity-preserving properties for motion by curvature. *Indiana Univ. Math. J.*, 65(1):171–197, 2016.
[87] Y. Lonke. Combinatorial problems in finite dimensional normed spaces (PhD dissertation, Hebrew university), 1998.
[88] Shachar Lovett and Raghu Meka. Constructive discrepancy minimization by walking on the edges. *SIAM Journal on Computing*, 44(5):1573–1582, 2015.
[89] Jiří Matoušek. *Geometric Discrepancy*. Springer Berlin, Heidelberg, 1999.
[90] Jiří Matoušek, Aleksandar Nikolov, and Kunal Talwar. Factorization norms and hereditary discrepancy. *International Mathematics Research Notices*, 2020(3):751–780, 02 2020.
[91] Emanuel Milman. On the role of convexity in isoperimetry, spectral gap and concentration. *Inventiones mathematicae*, 177:1 – 43, 2009.
[92] Aleksandar Nikolov, Kunal Talwar, and Li Zhang. The geometry of differential privacy: the sparse and approximate cases. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '13, page 351–360, New York, NY, USA, 2013. Association for Computing Machinery.
[93] Francesco Orabona and Dávid Pál. Optimal non-asymptotic lower bound on the minimax regret of learning with expert advice, 2015. Available at https://arxiv.org/abs/1511.02176.
[94] Stanley Osher and Ronald Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Applied Mathematical Sciences. Springer New York, NY, 2002.
[95] Yuval Peres, Oded Schramm, Scott Sheffield, and David Wilson. Tug-of-war and the infinity Laplacian. *Journal of the American Mathematical Society*, 22(1):167–210, 2009.
[96] Lucas Pesenti and Adrian Vladu. Discrepancy minimization via regularization. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1734–1758. SIAM, 2023.
[97] L. S. Pontryagin. The mathematical theory of optimal processes and differential games. volume 169, pages 119–158, 254–255. 1985. Topology, ordinary differential equations, dynamical systems.
[98] Thomas Rothvoss. Constructive discrepancy minimization for convex sets. *SIAM Journal on Computing*, 46(1):224–234, 2017.
[99] Luis Silvestre. Viscosity solutions of elliptic equations, 2015. Available at https://www.math.uchicago.edu/~luis/preprints/viscosity-solutions.pdf.
[100] Shridhar Sinha. The structure of extremal bad science matrices, 2025.
[101] Gleb Smirnov and Roman Vershynin. Discrepancy and Fisher information, 2026.
[102] H Mete Soner and Nizar Touzi. Dynamic programming for stochastic target problems and geometric flows. *Journal of the European Mathematical Society*, 4(3):201–236, 2002.
[103] H Mete Soner and Nizar Touzi. A stochastic representation for mean curvature type geometric flows. *The Annals of probability*, 31(3):1145–1165, 2003.
[104] Joel Spencer. Balancing games. *Journal of Combinatorial Theory, Series B*, 23(1):68–74, 1977.
[105] Joel Spencer. Six standard deviations suffice. *Transactions of the American Mathematical Society*, 289(2):679–706, 1985.
[106] Joel Spencer. Balancing vectors in the max norm. *Combinatorica*, 6(1):55–65, 1986.
[107] Joel Spencer. *Ten Lectures on the Probabilistic Method*. Society for Industrial and Applied Mathematics, 2nd edition edition, 1994.
[108] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463, May 2004.
[109] Stefan Steinerberger. Bad science matrices, 2024.
[110] Yoshihiro Tonegawa. *Brakke's Mean Curvature Flow: An Introduction*. Springer, 2019.
[111] Giancarlo Travaglini. *Number Theory, Fourier Analysis and Geometric Discrepancy*. London Mathematical Society Student Texts. Cambridge University Press, 2014.
[112] Paxton Turner, Raghu Meka, and Philippe Rigollet. Balancing Gaussian vectors in high dimension. In Jacob Abernethy and Shivani Agarwal, editors, *Proceedings of Thirty Third Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 3455–3486. PMLR, 09–12 Jul 2020.

Appendix A. Proofs of some standard propositions

In this Appendix, we record the proofs of several propositions used in the paper. These propositions are relatively standard and might distract from the main ideas in the paper, so their proofs have been relegated here. In what follows, we review the proofs of Proposition 5.3, Lemma 6.5, and also state and prove Proposition A.1.

Proof of Proposition 5.3. We are writing down the formula for the remainder term in a Taylor expansion, which we review for completeness and clarity on how the error term depends on ρ. Observe that

$$\partial_t^{-h}\phi(x,t) - \partial_t\phi(x,t) = \tfrac{1}{h}(\phi(x,t) - \phi(x,t-h) - h\partial_t\phi(x,t)).$$

Therefore, the first half of the proposition is equivalent to the statement

$$|\phi(x,t) - \phi(x,t-h) - h\partial_t \phi(x,t)| \leq \rho(h)h.$$

Indeed, integrating $\partial_t \phi$ we have

$$\begin{aligned}\phi(x,t) &= \phi(x,t-h) + \int_{t-h}^{t} \partial_s \phi(x,s) ds \\ &= \phi(x,t-h) + h\partial_t \phi(x,t) + \int_{t-h}^{t} ((\partial_s \phi)(x,s) - (\partial_s \phi)(x,t))\, ds.\end{aligned}$$

Since $\rho(\cdot)$ is a modulus of continuity for $\partial_t \phi$, the integral is on the right has a modulus not larger than $h\rho(h)$, from where the first bound follows. Next, we look at the spatial derivatives, let $z \in \mathbb{R}^m$, then

$$\begin{aligned}\nabla\phi(x+z,t) &= \nabla\phi(x,t) + \int_0^1 \frac{d}{ds}\nabla\phi(x+sz,t)\, ds \\ &= \nabla\phi(x,t) + \int_0^1 D^2\phi(x+sz,t)z\, ds \\ &= \nabla\phi(x,t) + D^2\phi(x,t)z + \int_0^1 (D^2\phi(x+sz,t) - D^2\phi(x,t))z\, ds.\end{aligned}$$

Integrating again, we have

$$\begin{aligned}&\phi(x+z,t) - \phi(x,t) - \nabla\phi(x,t)\cdot z \\ &= \int_0^1 (\nabla\phi(x+s_1 z,t) - \nabla\phi(x,t))\cdot z\, ds_1 \\ &= \int_0^1 (D^2\phi(x,t)z)\cdot z s_2\, ds_2 + \int_0^1\int_0^1 (D^2\phi(x+s_1s_2z,t) - D^2\phi(x,t))z\cdot z\, ds_1 ds_2.\end{aligned}$$

Rearranging, we have

$$\begin{aligned}&\phi(x+z,t) - \phi(x,t) - \nabla\phi(x,t)\cdot z - \tfrac{1}{2}D^2\phi(x,t)z\cdot z \\ &= \int_0^1\int_0^1 (D^2\phi(x+s_1s_2z,t) - D^2\phi(x,t))z\cdot z\, ds_1 ds_2.\end{aligned}$$

The integrand in the last double integral has absolute value no larger than $\omega(|z|)|z|^2$, and so

$$|\phi(x+z) - \phi(x) - z\cdot\nabla\phi(x) - \tfrac{1}{2}(D^2\phi(x)z,z))| \leq \rho(|z|)|z|^2.$$

We apply this inequality by taking $z = \delta A\bar{\varepsilon}$ and use it to bound the difference

$$\left|\max_A \min_{\bar{\varepsilon}} \delta^{-2}(\phi(x+\delta A\bar{\varepsilon}) - \phi(x)) - \tfrac{1}{2}\max_A \min_{\bar{\varepsilon}} \left\{2\delta^{-1}A\bar{\varepsilon}\cdot\nabla\phi(x) + (D^2\phi(x)A\bar{\varepsilon}, A\bar{\varepsilon})\right\}\right|,$$

by the maximum of $\rho(\delta\|A\bar{\varepsilon}\|_2)\|A\bar{\varepsilon}\|_2^2$ over all admissible A and $\bar{\varepsilon}$. As the difference is simply $\mathcal{H}_{\delta,n}(\phi)(x)$ minus $H_{\delta,n}(D^2\phi(x),\nabla\phi(x))$, we have

$$\left|\mathcal{H}_{\delta,n}(\phi)(x) - H_{\delta,n}(D^2\phi(x),\nabla\phi(x))\right| \leq 2\max_{A,\bar{\varepsilon}}\left\{\rho(\delta\|A\bar{\varepsilon}\|_2)\|A\bar{\varepsilon}\|_2^2\right\} \leq 2\rho(n\delta)n^2$$

□

Proof of Lemma 6.5. Consider first the case where we have a smooth function ϕ such that $w^* - \phi$ has a strict maximum at some (x_0,t_0). Then, it can be shown (see Proposition A.1 below there are sequences $\delta_k \to 0^+$ and (x_k,t_k) with $(x_k,t_k) \to (x_0,t_0)$ and $t_k \in \delta_k^2\mathbb{Z}_-$ such that $w_{\delta_k} - \phi$ has a local maximum at (x_k,t_k) for every k. In particular, we have the inequality

$$\partial_t^{-\delta_k^2}\phi(x_k,t_k+\delta^2) + \tfrac{1}{2}\mathcal{H}_{\delta_k,n}(\phi)(x_k,t_k+\delta^2) \geq 0. \tag{A.1}$$

Indeed, since $w_{\delta_k} - \phi$ has a local maximum at (x_k,t_k), we have

$$\phi(x',t') - \phi(x_k,t_k) \geq w_{\delta_k}(x',t') - w_{\delta_k}(x_k,t_k),$$

for any $x' \in \mathbb{R}^m$, $t' \in \delta_k^2\mathbb{Z}_-$. In particular, for every A and $\bar{\varepsilon}$ we have

$$\phi(x_k+\delta_k A\bar{\varepsilon}, t_k+\delta_k^2) - \phi(x_k,t_k) \geq w_{\delta_k}(x_k+\delta_k A\bar{\varepsilon}, t_k+\delta_k^2) - w_{\delta_k}(x_k,t_k).$$

Taking the minimum for all admissible $\overline{\varepsilon}$ followed by the maximum over all admissible A it follows that

$$\begin{aligned}&-\phi(x_k,t_k)+\max_A\min_{\overline{\varepsilon}}\phi(x_k+\delta_kA\overline{\varepsilon},t_k+\delta_k^2)\\&\geq -w_{\delta_k}(x_k,t_k)+\max_A\min_{\overline{\varepsilon}}w_{\delta_k}(x_k+\delta_kA\overline{\varepsilon},t_k+\delta_k^2)=0.\end{aligned}$$

Then, from the definition of $\partial_t^{-\delta_k^2}$ and $\mathcal{H}_{\delta_k,n}$ (see Section 5) we have

$$\partial_t^{-\delta_k^2}\phi(x_k,t_k+\delta_k^2)+\tfrac{1}{2}\mathcal{H}_{\delta_k,n}(\phi)(x_k,t_k+\delta_k^2)\geq 0,$$

and thus we obtain the inequality in (A.1). Now, with $\rho(\cdot)$ denoting a modulus of continuity for $\partial_t\phi$ and $D^2\phi$ we apply Proposition 5.3 and obtain

$$|\partial_t^{-\delta_k^2}\phi(x_k,t_k+\delta_k^2)-\partial_t\phi(x_k,t_k+\delta_k^2)|\leq\rho(\delta_k^2),$$

as well as

$$|\mathcal{H}_{\delta_k,n}(\phi)(x_k,t_k+\delta_k^2)-H_{\delta_k,n}(D^2\phi,\nabla\phi)(x_k,t_k+\delta_k^2)|\leq 2\rho(\delta_kn)n^2.$$

In particular,

$$\partial_t\phi(x_k,t_k+\delta_k^2)+\tfrac{1}{2}H_{\delta_k,n}(D^2\phi(x_k,t_k+\delta_k^2),\nabla\phi(x_k,t_k+\delta_k^2))\geq-\rho(\delta_k^2)-2\rho(\delta_kn)n^2.$$

We now use Lemma 5.4 which states that

$$\limsup_k H_{\delta_k,n}(M_k,p_k)\leq(H_n)^*(M,p),$$

and conclude that

$$\partial_t\phi(x_0,t_0)+\tfrac{1}{2}(H_n)^*(D^2\phi(x_0,t_0),\nabla\phi(x_0,t_0))\geq-\lim_{\delta\to0^+}\{\rho(\delta^2)+2\rho(\delta n)n^2\}=0.$$

This shows w^* is a viscosity subsolution. Now suppose ϕ is a smooth function such that $w_*-\phi$ has a local minimum at (x_0,t_0). As before, we can produce a sequence $\delta_k\to0^+$ and (x_k,t_k) with $t_k\in\delta_k^2\mathbb{Z}_-$ such that $w_{\delta_k}-\phi$ has a local minimum at (x_k,t_k), in which case

$$\partial_t^{-\delta_k^2}\phi(x_k,t_k+\delta_k^2)+\tfrac{1}{2}\mathcal{H}_{\delta_k,n}(\phi)(x_k,t_k+\delta_k^2)\leq 0.$$

Again as before, we use Proposition 5.3 and conclude that

$$\partial_t\phi(x_k,t_k+\delta_k^2)+\tfrac{1}{2}H_{\delta,n}(D^2\phi(x_k,t_k+\delta_k^2),\nabla\phi(x_k,t_k+\delta_k^2))\leq\rho(\delta_k^2)+2\rho(\delta_kn)n^2,$$

then, according to Lemma 5.4 we have

$$\liminf_k H_{\delta_k,n}(M_k,p_k)\geq(H_n)_*(M,p),$$

and we conclude that

$$\partial_t\phi(x_0,t_0)+\tfrac{1}{2}(H_n)_*(D^2\phi(x_0,t_0),\nabla\phi(x_0,t_0))\leq 0.$$

To finish the proof of the lemma it remains to consider the case where ϕ is such that $w^*-\phi$ has a maximum (or minimum) at some (x_0,t_0) that is not necessarily strict. We explain the procedure for the case of the maximum. In this case, we consider for $\beta>0$ and small the function

$$\phi_\beta(x,t)=\phi(x,t)+\beta(|x-x_0|^2+(t-t_0)^2).$$

Then the function $w^*-\phi_\beta=w^*-\phi-\beta(|x-x_0|^2+(t-t_0)^2)$ has a strict maximum at (x_0,t_0), and so by the previous discussion we have

$$\partial_t\phi_\beta(x_0,t_0)+\tfrac{1}{2}(H_n)^*(D^2\phi_\beta(x_0,t_0),\nabla\phi_\beta(x_0,t_0))\geq 0.$$

This holds for every $\beta>0$, and taking $\beta\to0$ we obtain the corresponding inequality for ϕ. The case where $w_*-\phi$ has a not-necessarily strict minimum is handled similarly, and this completes the proof of the lemma. □

The following proposition is a straightforward argument in real analysis, and it is a modification of a basic argument seen often in the viscosity solutions literature. It is used in Section 6 to show the viscosity subsolution/supersolution property for the half-relaxed limits of w_δ.

Proposition A.1. *Suppose $u_k : \mathbb{R}^m \times \delta_k^2\mathbb{Z}_- \to \mathbb{R}$ is a sequence of locally uniformly bounded functions continuous in x, and denote by u^* their upper semi-continuous limit. Then, if ϕ is a smooth function such that $u^* - \phi$ has a strict local maximum at $(x_{\max}, t_{\max})$ we can find a sequence $\{(x_k, t_k)\}$ with $(x_k, t_k) \in \mathbb{R}^m \times \delta_k^2\mathbb{Z}_-$ for every k such that $x_k \to x_{max}, t_k \to t_{max}$ and $u_k - \phi$ has a local maximum at (x_k, t_k).*

Remark A.2. *A respective version of the proposition holds if one instead considers the lower semi-continuous limit and a smooth function ϕ such that $u_* - \phi$ has a strict local minimum – one simply needs to apply the proposition above to the sequence $-u_k$.*

Proof. By assumption there exists $\epsilon_0 > 0$ so that $(x_{\max}, t_{\max})$ is the unique maximum of $u^* - \phi$ in

$$D := \{(x,t) : |x - x_{\max}|^2 + |t - t_{\max}|^2 \leq \epsilon_0^2\}.$$

Given the continuity of each u_k in x, there is a maximum point (x_k, t_k) for $u_k - \phi$ in the compact set[8]

$$D_k := \left\{(x,t) : |x - x_{\max}|^2 + |t - t_{\max}|^2 \leq \epsilon_0^2,\ t \in \delta_k^2\mathbb{Z}_-\right\}.$$

By compactness, the sequence $\{(x_k, t_k)\}$ has a non-empty set of limit points, and again by compactness the sequence will actually converge if there is exactly one limit point. Suppose (x_*, t_*) is such a limit point, we are going to show $(x_*, t_*) = (x_{\max}, t_{\max})$ and thus prove the proposition.

Consider a sequence $(\hat{x}_k, \hat{t}_k)$ such that $(\hat{x}_k, \hat{t}_k) \in D_k$, $(\hat{x}_k, \hat{t}_k) \to (x_{\max}, t_{\max})$ as $k \to \infty$, and

$$\lim_k u_k(\hat{x}_k, \hat{t}_k) = u^*(x_{\max}, t_{\max}).$$

Then, since (x_k, t_k) achieves the maximum of $u_k - \phi$ in D_k, we have

$$u_k(\hat{x}_k, \hat{t}_k) - \phi(\hat{x}_k, \hat{t}_k) \leq u_k(x_k, t_k) - \phi(x_k, t_k)\ \forall\, k.$$

Since (x_*, t_*) is a limit point of $\{(x_k, t_k)\}_k$ there is a sequence $j_k \to \infty$ as $k \to \infty$ such that $(x_{j_k}, t_{j_k}) \to (x_*, t_*)$. In this case, we may pass to the limit in the last inequality and conclude that

$$u^*(x_{\max}, t_{\max}) - \phi(x_{\max}, t_{\max}) \leq u^*(x_*, t_*) - \phi(x_*, t_*).$$

This means (x_*, t_*) achieves the maximum of $u^* - \phi$ in D, and so $(x_*, t_*) = (x_{\max}, t_{\max})$ as we wanted, finishing the proof.

□

Appendix B. Basics of viscosity solutions

In this Appendix, we recall some fundamental definitions and results from the viscosity solution theory for the equation

$$\partial_t w + \tfrac{1}{2} H_n(D^2 w, \nabla w) = 0 \text{ in } \mathbb{R}^m \times (-T, 0), \tag{B.1}$$

posed for some $T > 0$.

The strength of the viscosity solutions theory resides in the comparison principle. Roughly speaking, if u, v are two viscosity solutions and $u \leq v$ at $t = 0$, then $u \leq v$ for all times. This comparison property holds if u, v do not satisfy the PDE but rather satisfy respective partial differential inequalities – this is what brings us to the notion of subsolution and supersolution. Moreover, viscosity solutions deal with operators satisfying a monotonicity property (recall the "Global Comparison Property" discussed towards the end of Section 1.4), and it is this monotonicity that permits this notion of solution where one checks the PDE by "testing" against smooth functions that "touch" the subsolution/supersolution at a point.

In what follows, we will say ϕ is a *test function* if $\phi(x,t)$ is defined in some open set of $\mathbb{R}^m$ and the derivatives $\partial_t\phi$, $\nabla\phi(x,t)$, and $D^2\phi(x,t)$ are well defined and continuous in the domain of definition of ϕ.

Definition B.1. *An upper semi-continuous function $u : \mathbb{R}^m \times (-\infty, 0] \to \mathbb{R}$ is called a viscosity subsolution of (B.1) if for every test function $\phi(x,t)$ such that $u - \phi$ has a local maximum at some (x_0, t_0), $t_0 < 0$, we have*

$$\partial_t\phi(x_0, t_0) + \tfrac{1}{2}(H_n)^*(D^2\phi(x_0, t_0), \nabla\phi(x_0, t_0)) \geq 0.$$

[8] There are only finitely many possible values of t in this set, so one only need to take the largest maximum value among the finitely many time slices.

A lower semi-continuous function $v : \mathbb{R}^m \times (-\infty, 0] \to \mathbb{R}$ is called a viscosity supersolution of (B.1) *if for every smooth $\phi(x,t)$ such that $v - \phi$ has a local minimum at some (x_0, t_0), $t_0 < 0$, we have*

$$\partial_t \phi(x_0,t_0) + \tfrac{1}{2}(H_n)_*(D^2\phi(x_0,t_0), \nabla\phi(x_0,t_0)) \leq 0.$$

Finally, a function is called a viscosity solution of (B.1) *if it is both a viscosity subsolution and viscosity supersolution.*

We state the comparison principle for viscosity solutions of the equation. The proof follows largely along the lines for what is typically done in the literature, for instance the comparison principle in the first of the papers by Evans and Spruck, see [49, Theorem 3.2] and the more general comparison result in the work of Giga, Goto, Ishii, and Sato [56]. The theorem in [49] provides a comparison principle for the mean curvature flow equation, which corresponds to $n \geq m-1$ here. Strictly speaking, the comparison principle in [49] does not apply as stated to (B.1) when $n < m-1$. Here we rely on the more general comparison result in [56] which deals with a general geometric flow and allows us for terminal data that might grow at infinity (a must have if we are to work with a Banach space norms as terminal data). We state a special case of the comparison principle in [56] and briefly discuss how the assumptions are satisfied in the case of (B.1).

Theorem B.2. *Suppose u, v are respectively a viscosity subsolution and a viscosity supersolution of* (B.1) *for some $T > 0$, and suppose the following holds*

1. There is a $K > 0$ such that for all $x \in \mathbb{R}^m, t \in [-T, 0]$ we have

$$u(x,t) \leq K(|x|+1), v(x,t) \geq -K(|x|+1).$$

2. The functions $u(x,t), v(x,t)$ are continuous as $t \to 0$, that is, we have $u^*(x,0) = u_*(x,0)$ and $v^*(x,0) = v_*(x,0)$ for every $x \in \mathbb{R}^m$.

3. For some modulus of continuity ρ we have

$$u(x,0) - v(y,0) \leq \rho(|x-y|).$$

Then, we have

$$u(x,t) \leq v(x,t) \;\forall\, x \in \mathbb{R}^m,\; t \in [-T,0].$$

Proof. This is a special case of [56, Theorem 3.2], a result that deals with general unbounded domain U (here we only need it for $U = \mathbb{R}^m$) and a more general PDE $F(D^2u, \nabla u)$ (in our case, $F = \frac{1}{2}H_n$). The only requirements made on F in [56, Theorem 3.2] are

(1) $F(M,p)$ is continuous in the neighborhood of (M,p) when $p \neq 0$ and is degenerate elliptic in the sense that $F(M,p) \leq F(M+N,p)$ whenever N is positive semi-definite.
(2) We must have $F_*(0,0) = F^*(0,0)$.
(3) For every $R > 0$, there must be some finite constant $C_R > 0$ such that

$$|F(M,p)| \leq C_R \text{ whenever } |p| \leq R, \|M\|_{\mathrm{Fr}} \leq R.$$

Each of these properties are established $F = \frac{1}{2}H_n$ by the results at the end of Section 5, specifically Lemma 5.4 and Remark 5.5. □

As it is standard, the comparison principle immediately implies the uniqueness of viscosity solutions. It is also possible to use it to prove existence of such a solution via the Perron method (see, for instance [46, Section 4]). However, as solutions are built here from the continuum limit of the discrete game we will not discuss this method.

The comparison principle in Theorem B.2 is a key component in the proof of Theorem 6.1. These two theorems guarantee the existence and uniqueness of a viscosity solution to the terminal value problem (6.1). We also prove this unique solution is self-similar and has convex sublevel sets in space for each fixed time. The first of these results is made possible by the uniqueness of the solution to (6.1) (which follows from Theorem B.2) and the fact that the terminal data $x \mapsto \|x\|_E$ is a homogeneous function.

Lemma B.3. *If $w(x,t)$ is the unique viscosity solution of* (6.1)*, then for every $x \in \mathbb{R}^m$ and $t < 0$ we have*

$$w(x,t) = \sqrt{-t}\,w(x/\sqrt{-t}, -1).$$

Proof. From Theorem 6.1 we know that the function w, unique viscosity solution for (6.1), is the limit of w_δ as $\delta \to 0$. Therefore this function satisfies the pointwise bounds

$$\|x\|_E \leq w(x,t) \leq \sqrt{\|x\|_E^2 - Ct}.$$

Now, for $\lambda > 0$ define

$$w^{(\lambda)}(x,t) = \lambda^{-1} w(\lambda x, \lambda^2 t).$$

It follows from a straightforward computation and w being a viscosity solution of the PDE in (6.1) that $w^{(\lambda)}$ is a (viscosity) solution of the same PDE. On the other hand, from the homogeneity of the terminal data we have

$$w^{(\lambda)}(x,0) = \lambda^{-1}\|\lambda x\|_E = \|x\|_E = w(x,0).$$

That is, w and $w^{(\lambda)}$ are two viscosity solutions of the same terminal value problem. Moreover, these two functions satisfy the following two sided bounds for every x, t,

$$\|x\|_E \leq w(x,t),\ w^{(\lambda)}(x,t) \leq \sqrt{\|x\|_E^2 - Ct}.$$

In particular,

$$|w(x,t) - w^{(\lambda)}(x,t)| \leq \sqrt{C|t|}.$$

This means that $w, w^{(\lambda)}$ satisfy the growth assumptions in Theorem B.2. We conclude that

$$w(x,t) = w^{(\lambda)}(x,t),\ \forall\ (x,t) \in \mathbb{R}^m \times \mathbb{R}_-,$$

and this holds for any $\lambda > 0$. In particular, given x, t, choosing $\lambda = (\sqrt{-t})^{-1}$ we have

$$w(x,t) = \lambda^{-1} w(\lambda x, \lambda^2 t) = \sqrt{-t}\, w(x/\sqrt{-t}, -1),$$

and the lemma is proved. □

To finish this appendix, we discuss the convexity of the sublevel sets of u. For this, it will be convenient (as discussed in Section 5.2) to consider the terminal value problem

$$\begin{cases} \partial_t \overline{w} + \frac{1}{2}\overline{H}_n(D^2\overline{w}, \nabla\overline{w}) & = 0 \text{ in } \mathbb{R}^m \times \mathbb{R}_-, \\ \overline{w}(x,0) & = \|x\|_E \text{ in } \mathbb{R}^m. \end{cases} \tag{B.2}$$

The PDE of direct interest to us is (B.1), that of flow by the sum of principal positive curvatures. However, the PDE in (B.2) which is flow by the sum of principal curvatures (positive or not) is more standard and vastly studied in the literature. For solutions whose sublevel sets are convex the two equations agree, so one can drop the distinction of "positive" in the principal curvatures. For this reason it is an important fact that solutions to (B.2) always have convex sublevel sets in space for each time. We record this as a theorem, which we invoke in the proof of the main theorem at the end of Section 6.

Theorem B.4. *Consider $\overline{w} : \mathbb{R}^m \times (-T, 0] \to \mathbb{R}$ a viscosity solution of (B.2). Then for every $t \in (-T, 0]$ the sublevel sets of $\overline{w}(\cdot, t)$ are convex, in particular, $\overline{w}$ is also a viscosity solution of (B.1).*

We state this theorem without proof, and refer the reader to the relevant literature. First, we note that for $n \geq m - 1$ (so, the case of the standard mean curvature flow) this result is contained in Evans Spruck [49, Section 7.4], where they obtain the earlier landmark result of Huisken [69] but in the non-smooth viscosity solution setting (they do this by reducing the problem to one for an elliptic equation, from where the result for the level set function follows). For the other values of n (so, $n \leq m - 2$) the preservation of convexity result follows first by considering the work of Andrews [7] who considers more general functions of the principal curvatures in the smooth setting and shows the propagation of convexity, and then applying an approximation argument to pass his result to the flow by the sum of the largest k principal curvatures. Last but not least, it is worth mentioning a result of Qing, Schikorra, and Zhou [86] for the case $m = 2$ and $n = 1, 2$ notable in that they obtain a new proof of the preservation of convexity of sublevel sets for the mean curvature flow using not PDE but game theoretic tools, making use of the work of Kohn and Serfaty [78] (they also obtain analogous results for other flows using a game theoretic perspective).

Courant Institute, New York University, New York, NY 10012
Email address: nestor.g@nyu.edu

Department of Mathematics, The Ohio State University, Columbus, OH 43210
Email address: kobzar.1@osu.edu