The Uniformly Rotated Mondrian Kernel

Calvin Osborne
Harvard University

Eliza O'Reilly
Johns Hopkins University

Abstract

Random feature maps are used to decrease the computational cost of kernel machines in large-scale problems. The Mondrian kernel is one such example of a fast random feature approximation of the Laplace kernel, generated by a computationally efficient hierarchical random partition of the input space known as the Mondrian process. In this work, we study a variation of this random feature map by applying a uniform random rotation to the input space before running the Mondrian process to approximate a kernel that is invariant under rotations. We obtain a closed-form expression for the isotropic kernel that is approximated, as well as a uniform convergence rate of the uniformly rotated Mondrian kernel to this limit. To this end, we utilize techniques from the theory of stationary random tessellations in stochastic geometry and prove a new result on the geometry of the typical cell of the superposition of uniformly rotated Mondrian tessellations. Finally, we test the empirical performance of this random feature map on both synthetic and real-world datasets, demonstrating its improved performance over the Mondrian kernel on a dataset that is debiased from the standard coordinate axes.

1 INTRODUCTION

Random feature kernel approximations were introduced by Rahimi and Recht (2008) to mitigate the high computational cost of kernel methods in large-scale problems. Instead of using a kernel to implicitly lift data using some map ϕ into an infinite dimensional feature space, they proposed an explicit embedding of the data using a low-dimensional *random* feature map z such that the inner product in this feature space approximates the kernel evaluation:

$$K(x, x') = \langle \phi(x), \phi(x') \rangle \approx z(x)^T z(x').$$

Rahimi and Recht (2008) proposed two different random feature maps in their work: *Fourier random features*, which are obtained using Bochner's theorem for stationary kernels, and *random binning features*, which are obtained by partitioning the input space with random axis-aligned hyperplanes and generating a feature map indicating the cell of the partition the input is contained in. The former has received much more attention in subsequent literature, see the recent survey by Liu et al. (2022) and the references therein. However, it has been observed by Wu et al. (2016) that random binning features have computational advantages and improved performance over Fourier random features for several tasks.

Another random feature kernel generated using random partitions of the input space was proposed by Balog et al. (2016) and called the *Mondrian kernel*. This kernel is constructed by using the random binning feature map induced by a random hierarchical partition called the *Mondrian process* (Balog and Teh, 2015) that has many appealing properties such as the Markov property and spatial consistency. This partition produces a feature map approximating the Laplace kernel, the same kernel that is approximated by the random binning feature map (Rahimi and Recht, 2008). However, the special properties of the process provide an efficient bandwidth learning procedure (Balog et al., 2016), a vital parameter tuning step that requires expensive cross-validation procedures in the case of kernel methods, Fourier random features, and random binning features. In addition, Mondrian forests and other oblique variants have been shown to achieve minimax-optimal convergence rates in regression problems (Mourtada et al., 2020; Cattaneo et al., 2023; O'Reilly and Tran, 2021).

However, one drawback of the Mondrian kernel and random binning features kernel is that the expressiveness of these random feature maps is limited since the

Proceedings of the 28[th] International Conference on Artificial Intelligence and Statistics (AISTATS) 2025, Mai Khao, Thailand. PMLR: Volume 258. Copyright 2025 by the author(s).

partitions are restricted to axis-aligned splits. Indeed, as mentioned above, the Mondrian kernel and the random binning features kernel studied in Wu et al. (2016) are both known to approximate the Laplace kernel, whereas Fourier random features have the power to approximate *any* stationary, positive definite, symmetric kernel. Recent work by O'Reilly and Tran (2022) has shown that a much larger class of kernels can be approximated by using oblique random partitions, increasing the flexibility of this random feature map. In particular, this work studied the random binning feature map generated by stable under iteration (STIT) processes (Nagel and Weiss, 2005) in stochastic geometry. These stochastic processes are quite expressive, as they are defined for any dimension d and are indexed by a probability measure on the unit sphere S^{d-1}, called the *directional distribution*, that contains d linearly independent unit vectors in its support. For example, it was observed that the translation and rotation invariant exponential kernel could be approximated when this directional distribution is the uniform measure on the unit sphere. This class of STIT processes includes the Mondrian process as a particular case when the directional distribution is the uniform measure on the standard basis vectors.

Learning methods utilizing random partitions with oblique splits have been shown to obtain improved empirical performance over axis-aligned versions for many datasets (Blaser and Fryzlewicz, 2016; Ge et al., 2019). However, a major hurdle in adopting these approaches is the increased computational costs in generating splits that depend on linear combinations of up to all d dimensions on the input instead of just one dimension in the axis-aligned case. For example, it was shown by Ge et al. (2019) that isotropic STIT processes generate random forests that obtain improved empirical performance over Mondrian forests, but there are significant computational difficulties in generating an isotropic STIT process rather than a Mondrian process.

O'Reilly and Tran (2022) showed that any STIT process with a discrete directional distribution can be obtained by lifting to a higher dimensional space and running a Mondrian process. However, the induced class of kernels does not include an isotropic one. This work proposes a variant of the Mondrian kernel that approximates a kernel that is invariant under rotation without implementing a complex isotropic STIT process. In particular, we study the random feature map generated by random partitions with oblique splits by applying a uniformly random rotation to the input space and then partitioning the transformed space with a Mondrian process. We call the associated random feature kernel the *uniformly rotated Mondrian kernel*.

Our theoretical contributions in this work include characterizing the *isotropic* kernel that is approximated by the uniformly rotated Mondrian kernel as the number of random features approaches infinity, as well as a uniform rate of convergence to this limiting kernel. Similar to the results in O'Reilly and Tran (2022), these findings will depend on machinery from the theory of stationary random tessellations in stochastic geometry (Schneider and Weil, 2008), as the superposition of uniformly rotated Mondrian tessellations is composed of cells that are more general polytopes than axis-aligned boxes.

Beyond these theoretical contributions, we empirically study the performance of the uniformly rotated Mondrian kernel on synthetic and real-world data sets. To evaluate the performance of this random feature kernel in comparison to Fourier random features, random binning features, and the Mondrian kernel, we first run comparable experiments on the same datasets that were studied in the proposal of the Mondrian kernel in Balog et al. (2016). These experiments show that the uniformly rotated Mondrian kernel can achieve similar performance to the Mondrian kernel on what we will show to be an *adversarial* `CPU` dataset without sacrificing computational efficiency. We then evaluate the performance of the uniformly rotated Mondrian kernel and the Mondrian kernel on a non-adversarial dataset, the *Mondrian line*, to show that the uniformly rotated Mondrian kernel can indeed outperform the Mondrian kernel on a dataset that is debiased from a small number of coordinate axes. In particular, these experiments demonstrate that the uniformly rotated Mondrian kernel exhibits the improved empirical performance observed in other random partition kernels with oblique splits (Blaser and Fryzlewicz, 2016; Ge et al., 2019) without an expensive increase in computational demands as with the isotropic STIT process.

1.1 Related Work

Methods that utilize oblique splits have shown improved empirical performance across many tasks in machine learning, but general polyhedral cells incur increased computational and theoretical difficulties. Multiple works have proposed generating a partition by randomly rotating the feature space to mitigate this cost and then partitioning with axis-aligned splits. For example, random rotations have been used to increase the diversity of estimators in ensemble methods (Blaser and Fryzlewicz, 2016) and show improved performance over axis-aligned random forests. Randomly rotated k-d trees were studied by Vempala (2012) for nearest neighbor search and were shown to adapt to the intrinsic dimension of the input data. Our work studies this kind of approach when generating random

features with a Mondrian process. This allows us to obtain a closed-form expression of the limiting kernel, and convergence guarantees due to the connection with random tessellation models in stochastic geometry.

2 BACKGROUND

First, we will detail some background on random tessellations and, in particular, the Mondrian tessellation.

A *tessellation* X is a locally finite, countable collection of nonempty compact, convex subsets of $\mathbb{R}^d$ such that

$$\bigcup_{K \in X} K = \mathbb{R}^d$$

and $\operatorname{int} K \cap \operatorname{int} K' = \emptyset$ for all distinct pairs $K, K' \in X$. A *random tessellation* X is a point process on the space of nonempty compact subsets of $\mathbb{R}^d$ such that X is almost surely a tessellation.

Of particular interest to us is the *Mondrian tessellation of lifetime* $\lambda > 0$, a random tessellation constructed on an axis-aligned box $\mathcal{X} \subset \mathbb{R}^d$ by the following *Mondrian process*:

1. For each $1 \leq n \leq d$, we sample $t_n \sim \text{Exp}(|\mathcal{X}_n|)$, where $|\mathcal{X}_n|$ is the length of the box in the n-th dimension, and let $n_{\min} = \text{argmin}_n(t_n)$.
2. If $t_{n_{\min}} \geq \lambda$, then the construction is finished.
3. Otherwise, we sample $a \sim \text{Unif}(|\mathcal{X}_{n_{\min}}|)$ and split $\mathcal{X}$ into boxes $\mathcal{X}^{<} = \{x \in \mathcal{X} \mid x_{n_{\min}} \leq a\}$ and $\mathcal{X}^{>} = \{x \in \mathcal{X} \mid x_{n_{\min}} \geq a\}$.
4. We run this same process recursively and independently on $\mathcal{X}^{<}$ and $\mathcal{X}^{>}$ with lifetime $\lambda - t_{n_{\min}} > 0$.

See Figure 1 for two particular samples of a Mondrian tessellation on $\mathcal{X} = [0,1]^2$ with lifetime $\lambda = 1$.

The second class of random tessellations that will be of interest to us are *stationary Poisson hyperplane tessellations*, a class of random tessellations defined by a stationary Poisson process on the space of hyperplanes in $\mathbb{R}^d$ that constitute the boundaries of their cells. The distribution of these tessellations is determined by a constant intensity $\lambda > 0$ and a directional distribution φ on the unit sphere governing the distribution of the normal vectors of the hyperplanes. If φ is the uniform distribution on the unit basis vectors in $\mathbb{R}^d$, the random tessellation is called a *Poisson Manhattan tessellation*. See Figure 2 for two particular samples of a Poisson hyperplane tessellation on $\mathcal{X} = [0,1]^2$ with intensity $\lambda = 1$, one with a directional distribution that is uniform on the unit sphere, i.e., an *isotropic Poisson hyperplane tessellation*, and the other a Poisson Manhattan tessellation.

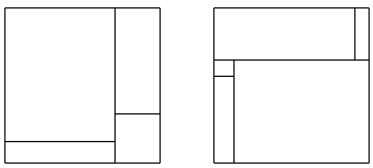


Figure 1: Sample Mondrian Tessellations on $\mathcal{X} = [0,1]^2$ with $\lambda = 1$.

3 UNIFORMLY ROTATED MONDRIAN KERNEL

We will now define the *random feature map* z for the uniformly rotated Mondrian process on a dataset $\{x_n\}_{1 \leq n \leq N}$ with $x_n \in \mathbb{R}^d$ as follows.

1. First, sample a random rotation $R \sim \text{Unif}(SO_d)$, and map each $x_n \mapsto Rx_n$.
2. Then, let $\mathcal{X}$ be a bounding box for the rotated dataset, and construct a Mondrian tessellation X on $\mathcal{X}$. Assign to each cell of X an index $1, 2, \ldots$.
3. Finally, define $z(x)$ to be the one-hot encoding of the index of the cell containing Rx, i.e., $z(x) \in \mathbb{R}^C$ where C is the number of cells in X, and $z(x)$ has a single coordinate value of one and zero elsewhere.

With this construction of a random feature map $z : \mathbb{R}^d \to \mathbb{R}^C$, we compute the *kernel map* $k : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ as

$$k(x, x') = z(x)^T z(x') = \begin{cases} 1 & \text{if } Rx, Rx' \text{ are in the same cell of } X, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, we extend this definition by generating M uniformly rotated Mondrian kernel maps $k^{(1)}, \ldots k^{(M)}$ induced by M independently generated uniform rotations $R^{(1)}, \ldots, R^{(M)}$ and Mondrian tessellations

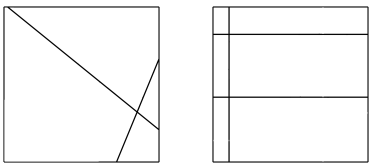


Figure 2: Sample Isotropic Poisson Hyperplane Tessellation and Poisson Manhattan Tessellation on $\mathcal{X} = [0,1]^2$ with $\lambda = 1$.

$X^{(1)}, \ldots, X^{(M)}$, and we define $k_M : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ to be

$$k_M(x, x') = \frac{1}{M} \sum_{m=1}^{M} k^{(m)}(x, x').$$

We call k_M the *uniformly rotated Mondrian kernel of order M*. By the law of large numbers, the sequence of uniformly rotated Mondrian kernels k_M of order M converges almost surely as $M \to \infty$ to the limiting kernel k_∞ given by

$$\begin{aligned} k_\infty(x, x') &= \mathbb{E}[k(x, x')] \\ &= \mathbb{P}[Rx, Rx' \text{ in the same cell of } X], \end{aligned} \quad (1)$$

i.e., given by the probability that the inputs are in the same cell of a uniformly rotated Mondrian tessellation.

4 MAIN RESULTS

We will now detail and discuss the two main theoretical results in this paper, with proofs to follow below.

4.1 Limiting Uniformly Rotated Mondrian Kernel

Our first result is an expression for the limiting kernel k_∞ of the uniformly rotated Mondrian process as is defined in (1).

Theorem 1. *The limiting kernel of the uniformly rotated Mondrian process with lifetime $\lambda > 0$ in $\mathbb{R}^d$ is of the form*

$$k_\infty(x, x') = \frac{1}{\omega_d} \int_{S^{d-1}} e^{-\lambda \|x - x'\|_2 \|v\|_1} \mathrm{d}v, \quad (2)$$

where ω_d is the surface area of the unit ball in $\mathbb{R}^d$.

We note that (2) is indeed isotropic since it depends only on $\|x - x'\|_2$, unlike the limiting Laplace kernel of the standard Mondrian process in Balog et al. (2016) which depends only on $\|x - x'\|_1$. See Figure 3 for a plot of this kernel in $\mathbb{R}^2$ with lifetime $\lambda = 1$. As will be detailed in Section 6, this result follows from the computation of an integral over the special orthogonal group SO_d with respect to the normalized Haar measure μ on SO_d.

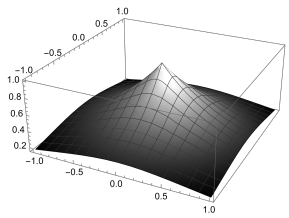


Figure 3: Limiting Kernel for the Uniformly Rotated Mondrian Process on $\mathbb{R}^2$ with lifetime $\lambda = 1$.

4.2 Uniform Convergence of Uniformly Rotated Mondrian Kernel

Our second result characterizes the rate of convergence of the sequence k_M to k_∞ as $M \to \infty$.

Theorem 2. *For any bounded domain $\mathcal{X} \subset \mathbb{R}^d$ and small enough $\delta > 0$, we have that*

$$\begin{aligned} \mathbb{P}\left[\sup_{x, x' \in \mathcal{X}} |k_M(x, x') - k_\infty(x, x')| > \delta\right] \\ \in \mathcal{O}\left(M^{d + d/(2d+1)} e^{-M\delta^2/(4d+2)}\right), \end{aligned} \quad (3)$$

where $k_M : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is the uniformly rotated Mondrian kernel of order M.

Similar to Proposition 2 in Balog et al. (2016), this second result guarantees the exponential uniform convergence of k_M to k_∞. Furthermore, we note that this rate of convergence is identical to that of the class of random feature kernels generated by STIT tessellations from Theorem 4.3 in O'Reilly and Tran (2022).

5 CELLS OF UNIFORMLY ROTATED MONDRIAN TESSELLATIONS

To prove Theorem 2, we will need to study the geometry of the cells of the random tessellation of the input space generated by the rotations and Mondrian partitions that constitute the uniformly rotated Mondrian kernel. In particular, we will study the geometry of the *typical cell*. The typical cell Z of a stationary random tessellation is a random nonempty compact, convex polytope with distribution as in (4.8) of Schneider and Weil (2008) given by choosing a cell uniformly on a compact, convex window $rW \subset \mathbb{R}^d$ and letting $r \to \infty$. In this section, we will state three results concerning the volume, inradius, and circumradius of the typical cell. Then, to prove Theorem 2, we will make frequent use of an application of Campbell's theorem as follows.

Theorem 3. (Schneider and Weil, 2008, Equation 4.3) *For any nonnegative measurable function f on the space of nonempty compact sets in $\mathbb{R}^d$, and for any stationary random tessellation X, we have that*

$$\mathbb{E}\left[\sum_{C \in cells(X)} f(C)\right] = \frac{1}{\mathbb{E}[V(Z)]} \cdot \mathbb{E}\left[\int_{\mathbb{R}^d} f(Z + y) \mathrm{d}y\right],$$

where Z is the typical cell of the X.

Conditioned on the uniform rotation R, the random tessellation of the input space generated by a uniformly rotated Mondrian kernel of order one is what

we will call a *rotated Mondrian tessellation*, that is, a STIT tessellation with directional distribution

$$\varphi = \frac{1}{d}\sum_{i=1}^{d}\delta_{u_i}, \tag{4}$$

where $\{u_1, \dots, u_d\}$ is an orthonormal set in $\mathbb{R}^d$ such that $u_i = Re_i$. We will then make use of Theorem 1 from Schreiber and Thäle (2013) which gives us that the typical cell of a rotated Mondrian tessellation is the same in distribution as the typical cell of a Poisson Manhattan tessellation with directional distribution (4) and the same lifetime or intensity parameter λ. We will call the latter a *rotated Poisson Manhattan tessellation*.

To now understand the asymptotics of the uniformly rotated Mondrian kernel of order M, we will need to study the *superposition* of independent uniformly rotated Mondrian tessellations. For any two tessellations X, X', their superposition is the tessellation composed of the pairwise intersection of their compact, convex cells. In the following, we will make use of a result from Page 158 of Schneider and Weil (2008) that states that the superposition of independent stationary Poisson processes is itself a stationary Poisson process, with an intensity measure given by the sum of the two component intensity measures. In summary, we will study the typical cell of the superposition of independent uniformly rotated Mondrian tessellations by considering the associated superposition of independent uniformly rotated Poisson Manhattan tessellations, which itself is a stationary hyperplane tessellation generated by a doubly stochastic Poisson hyperplane process.

With this background in mind, the results that we will establish to prove Theorem 2 are stated below.

Lemma 1. *Let X be the superposition of M independent rotated Poisson Manhattan tessellations $X_1, \dots, X_M$ of intensity $\lambda > 0$ in $\mathbb{R}^d$, and let Z be the typical cell of X. First, it holds that*

$$\frac{1}{\kappa_d}\left(\frac{2\sqrt{d}}{\lambda M}\right)^d \leq \mathbb{E}\left[V(Z)\right] \leq \frac{1}{\kappa_d}\left(\frac{2d}{M\lambda}\right)^d, \tag{5}$$

where $V(C)$ is the volume of a compact, convex body $C \subset \mathbb{R}^d$ and κ_d is the volume of the unit ball in $\mathbb{R}^d$. Second, it holds that

$$\mathbb{E}[r(Z)] = \frac{1}{2M\lambda}, \tag{6}$$

where $r(C)$ is the inradius of a compact, convex body $C \subset \mathbb{R}^d$. Finally, it holds that

$$\mathbb{P}[R(Z) \geq a] \leq e^{-2M\lambda a/d}\left(\sum_{n=0}^{d-1}\frac{1}{n!}\left[\frac{2\lambda a}{d}\right]^n\right)^M, \tag{7}$$

where $R(C)$ is the circumradius of a compact, convex body $C \subset \mathbb{R}^d$.

We will briefly remark here on the methods used to prove Lemma 1 and will provide the full details in the Appendix.

Equation (5) makes use of an object known as the *associated zonoid* Π_X of a Poisson hyperplane tessellation X. The associated zonoid Π_X is an explicit compact, convex body that is associated with the general shape of the cells in X. In particular, we rely on Theorem 10.3.3 in Schneider and Weil (2008) that states that $\mathbb{E}[V(Z)] = V(\Pi_X)^{-1}$.

Equation (6) follows from the more general Theorem 10.4.8 in Schneider and Weil (2008) that classifies the exact distribution of the inradius of the typical cell of a general Poisson hyperplane tessellation. These results can be applied to our class of superpositions of rotated Mondrian tessellations by Theorem 1 of Schreiber and Thäle (2013) as was discussed above.

Equation (7), to the best of our knowledge, makes use of a new technique to provide an explicit bound on the cumulative distribution function of the circumradius of the typical cell of the superposition of rotated Poisson hyperplane tessellations. One comparable result from Hug and Schneider (2007), applied by O'Reilly and Tran (2022) to STIT tessellations, gives a bound of the form

$$\mathbb{P}[R(Z) \geq a] \leq ce^{-\tau a h_{\min}(\Pi_X)}$$

for constants $c, \tau > 0$, where $h_{\min}(\Pi_X)$ is the minimum of the support function of the associated zonoid Π_X of the STIT tessellation X. However, since we do not know the dependence of the constant $c > 0$ on the number M of superpositions and the directional distribution φ_n of each rotated Mondrian tessellation, this result is not suitable for understanding the limiting behavior of the uniformly rotated Mondrian kernel as $M \to \infty$. As detailed in Proposition 1, our technique of lifting the tessellation X in $\mathbb{R}^d$ to the intersection of a higher-dimensional tessellation $\widetilde{X}$ in $\mathbb{R}^{Md}$ with a d-dimensional linear subspace U allows us to compute explicit constants and obtain (7).

6 PROOFS OF MAIN RESULTS

We will now discuss the proofs of Theorem 1 and Theorem 2, with further necessary details to be provided in the Appendix.

Proof of Theorem 1. We first compute the limiting kernel of a Mondrian process under a fixed rotation $R \in SO_d$, and we then compute the limiting kernel of the uniformly rotated Mondrian process by conditioning on the rotation $R \in SO_d$.

For a fixed rotation $R \in SO_d$, letting k_∞ be the rotated Mondrian limiting kernel and $\hat{k}_\infty$ be the Mondrian limiting kernel, we have from Proposition 1 in Balog et al. (2016) that

$$k_\infty(x, x') = \hat{k}_\infty(Rx, Rx') = e^{-\lambda||R(x-x')||_1}.$$

Letting μ be the normalized Haar measure on SO_d, we compute that

$$\begin{aligned} k_\infty(x, x') &= \mathbb{P}[k(x, x') = 1] \\ &= \int_{SO_d} \mathbb{P}[\hat{k}(Rx, Rx') = 1] \mathrm{d}\mu(R) \\ &= \int_{SO_d} \hat{k}_\infty(Rx, Rx') \mathrm{d}\mu(R) \\ &= \int_{SO_d} e^{-\lambda||R(x-x')||_1} \mathrm{d}\mu(R) \end{aligned}$$

from the law of total probability. Finally, from the translation invariance of the Haar measure μ, we have from right multiplication by the rotation $S \in SO_d$ that points $x - x'$ in the direction of the axis-aligned unit vector $e_1 \in \mathbb{R}^d$ that

$$\begin{aligned} k_\infty(x, x') &= \int_{SO_d} e^{-\lambda||RS(x-x')||_1} \mathrm{d}\mu(R) \\ &= \int_{SO_d} e^{-\lambda||x-x'||_2||Re_1||_1} \mathrm{d}\mu(R) \\ &= \frac{1}{\omega_d} \int_{S^{d-1}} e^{-\lambda||x-x'||_2||v||_1} \mathrm{d}v \end{aligned}$$

as is desired, completing our computation of the limiting kernel of the uniformly rotated Mondrian. □

We will now provide an overview of the proof of Theorem 2.

Outline of Proof of Theorem 2. We will first let B_r be a ball of radius r in $\mathbb{R}^d$ such that $\mathcal{X} \subseteq B_r$. We will then let $\mathcal{U}$ be an ε-grid covering of a concentric $B_{R+\varepsilon}$ for some $R > r$ and $\varepsilon > 0$ to be specified later. Letting X be the superposition of M independently rotated Mondrian tessellations, we will consider the following four events:

1. $A_1 = \{$every cell of X with a circumcenter outside of B_R is disjoint from $B_r\}$,
2. $A_2 = \{$every cell of X with a circumcenter in B_R contains a point of $\mathcal{U}\}$,
3. $A_3 = \{$every cell of X with a circumcenter in B_R has a diameter less than or equal to $\delta/4\lambda\sqrt{d}\}$,
4. $A_4 = \{$the $\delta/2$-approximation holds on $\mathcal{U}\}$,

where the *δ-approximation* on an ε-grid covering $\mathcal{U}$ is that

$$|k_M(u, u') - k_\infty(u, u')| \leq \delta$$

for all $u, u' \in \mathcal{U}$. It holds (see the Appendix) that $A_1 \cap A_2 \cap A_3 \cap A_4$ implies that the δ-approximation holds on $\mathcal{X}$, and so we have that

$$\begin{aligned} &\mathbb{P}\left[\sup_{x,x' \in \mathcal{X}} |k_M(x, x') - k_\infty(x, x')| > \delta\right] \\ &\leq \mathbb{P}(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c) \\ &\leq \mathbb{P}(A_1^c) + \mathbb{P}(A_2^c) + \mathbb{P}(A_3^c) + \mathbb{P}(A_4^c) \end{aligned}$$

from a union bound. It remains to bound the probability of the four events $A_1^c, A_2^c, A_3^c, A_4^c$. For A_1^c, letting $c(C)$ be the circumcenter of a nonempty compact, convex set, we note that

$$\begin{aligned} \mathbb{P}(A_1^c) &= \mathbb{P}\left[\bigcup_{\text{cells } C \in X} (c(C) \notin B_r) \cap (C \cap B_r \neq \emptyset)\right] \\ &\leq \mathbb{E}\left[\sum_{\text{cells } C \in X} \mathbb{1}[c(C) \notin B_r] \mathbb{1}[C \cap B_r \neq \emptyset]\right]. \end{aligned}$$

By Campbell's theorem and Lemma 1, we have that

$$\begin{aligned} \mathbb{P}(A_1^c) &\leq \left(\frac{1}{\mathbb{E}[V(Z)]}\right) \int_{||y||_2 \geq R} \mathbb{P}[(Z + y) \cap B_r \neq \emptyset] \mathrm{d}y \\ &\leq \left(\frac{1}{\mathbb{E}[V(Z)]}\right) \int_{||y||_2 \geq R} \mathbb{P}[r(Z) \geq ||y||_2 - r] \mathrm{d}y \\ &\leq \left(\frac{1}{\mathbb{E}[V(Z)]}\right) \int_{||y||_2 \geq R} e^{-2M\lambda(y-r)} \mathrm{d}y \\ &= \left(\frac{\kappa_d^2 d}{2\lambda M}\right) \left(\frac{\lambda M}{2\sqrt{d}}\right)^d e^{-2\lambda M(R-r)}, \end{aligned}$$

where Z is the typical cell of X and κ_d is the volume of the unit ball in $\mathbb{R}^d$. Deriving similar bounds for the other events, we find that

$$\begin{aligned} &\mathbb{P}\left[\sup_{x,x' \in \mathcal{X}} |k_M(x, x') - k_\infty(x, x')| > \delta\right] \\ &< \left(\frac{\kappa_d^2 d}{2\lambda M}\right) \left(\frac{\lambda M}{2\sqrt{d}}\right)^d e^{-2\lambda M(R-r)} \\ &+ \kappa_d^2 \left(\frac{\lambda R M}{2\sqrt{d}}\right)^d M\lambda\varepsilon\sqrt{d} \end{aligned}$$

$$+ \kappa_d^2 \left(\frac{\lambda R M}{2\sqrt{d}} \right)^d \left(\sum_{n=0}^{d-1} \frac{1}{n!} \left[\frac{\delta}{4d^{3/2}} \right]^n \right)^M e^{-M\delta/4d^{3/2}}$$

$$+ 2\varepsilon^{-2d} \left(\sum_{n=0}^{2d} \binom{2d}{n} (4R)^n \right) e^{-M\delta^2/2}.$$

This expression, when minimized with respect to ε, gives us the desired result for small enough $\delta > 0$. □

A detailed proof of Theorem 2 is included in the Appendix.

7 EXPERIMENTS

To evaluate the empirical performance of the uniformly rotated Mondrian kernel, we ran five experiments, the first four of which are similar to those in Balog et al. (2016), assessing the uniformly rotated Mondrian kernel against Fourier random features, random binning features, and the Mondrian kernel.

7.1 Experimental Procedures

In Experiment 1, we demonstrate the convergence of each random feature map to its limiting kernel, namely, the Laplace kernel for Fourier random features, random binning features, and the Mondrian process, and the limiting kernel identified in Theorem 1 for the uniformly rotated Mondrian process. To do so, we uniformly generated $n = 100$ points in $[0, 1]^2$ and plot the maximal absolute error between each random feature map and its limiting kernel over every pair of points for up to $M_{\max} = 50$ nonzero random features. For each random feature model, the lifetime parameter was set to $\lambda = 10$, and the experiment was repeated $\varepsilon = 5$ times to generate error bars.

In Experiment 2, we demonstrate, as in Balog et al. (2016), how the uniformly rotated Mondrian kernel can recover the ground truth lifetime of a synthetic dataset under an efficient kernel width selection procedure. To do so, we generate $N_{\text{train}} = N_{\text{test}} = N_{\text{validation}} = 500$ synthetic data points in $\mathbb{R}^2$ from the limiting kernel of the uniformly rotated Mondrian kernel with a lifetime parameter of $\lambda_0 = 10$. We then plot the train, test, and validation set errors from the uniformly rotated Mondrian kernel with $M = 50$ nonzero random features up to a maximum lifetime of $\lambda_{\max} = 30$. We recover at the minimum validation set error a lifetime of $\hat{\lambda} \approx 21$, which is comparable to the recovered lifetime of $\hat{\lambda} \approx 19$ from an similar experiment on the Mondrian kernel from Balog et al. (2016) with the same ground truth lifetime.

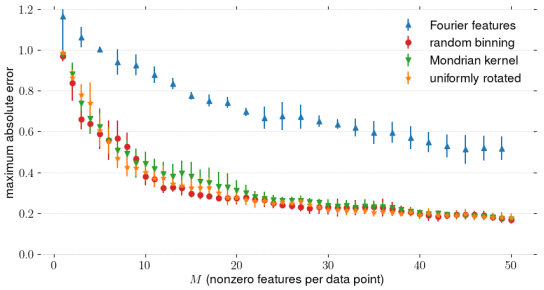
In Experiment 3, we compare the validation set errors of the random feature methods on the `CPU` dataset with respect to the number of nonzero random features, where, as in Balog et al. (2016), the primal optimization problems were solved by stochastic gradient descent with a Ridge regularization constant of $\delta^2 = 10^{-4}$. To do so, for each random feature method, the validation set error was computed for up to $M_{\max} = 50$ nonzero random features, and the experiment was repeated $\varepsilon = 5$ times to generate error bars. For Fourier random features, random binning features, and the Mondrian process, the lifetime parameter was set to $\lambda = 10^{-6}$ as in Balog et al. (2016), and for the uniformly rotated Mondrian process, the lifetime parameter was set to $\lambda = 2.5 \times 10^{-7}$ based on the results of a parameter sweep.

In Experiment 4, we compare the validation set errors of the random feature methods on the `CPU` dataset with respect to the total computational time, where we again solve the primal optimization problems as in Balog et al. (2016). To do so, for each random feature method, we used $M = 350$ nonzero random features. Furthermore, to set the lifetime parameter, a binary search algorithm as described in Balog et al. (2016) was implemented for Fourier random features and random binning features, while an efficient parameter sweep was used for the Mondrian kernel and uniformly rotated Mondrian kernel.

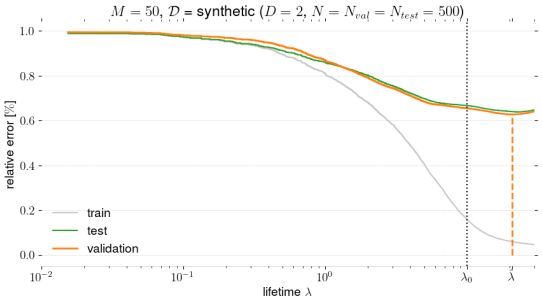
In Experiment 5, we compare the relative training and testing set error of the Mondrian kernel and uniformly rotated Mondrian kernel on a synthetic dataset we call the *Mondrian line*, which is adapted from the Mondrian cube in Ge et al. (2019). To generate the Mondrian line, we first sample $N_{\text{train}} = N_{\text{test}} = 500$ datapoints uniformly on $[0, 1] \times [-\varepsilon, \varepsilon] \times [-\varepsilon, \varepsilon] \subset \mathbb{R}^3$ for a small value of $\varepsilon = 0.01$. Then, we label a datapoint x_n as 1 if either $x_n \in [0, 1] \times [0, \varepsilon] \times [0, \varepsilon]$ or $x_n \in [0, 1] \times [-\varepsilon, 0] \times [-\varepsilon, 0]$ and 0 otherwise. Finally, we rotate each datapoint so that the dataset is biased toward the axis $\text{Span}(1, 1, 1)$. After constructing the dataset, we run Ridge regression with both the Mondrian kernel and uniformly rotated Mondrian kernel on the Mondrian line with $M = 500$ nonzero random features and plot the relative error with respect to the current lifetime.

7.2 Discussion of Results

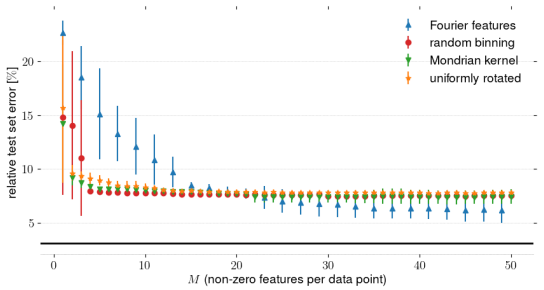
In Experiment 1, we confirm that the uniformly rotated Mondrian kernel indeed converges to an isotropic kernel, and, moreover, that it does so at a rate comparable to both random binning features and the Mondrian kernel, as well as faster than random Fourier features. In Experiment 2, we confirm that the uniformly rotated Mondrian kernel maintains an essential feature of the Mondrian kernel, namely, the fast recovery of a ground truth lifetime. This allows the



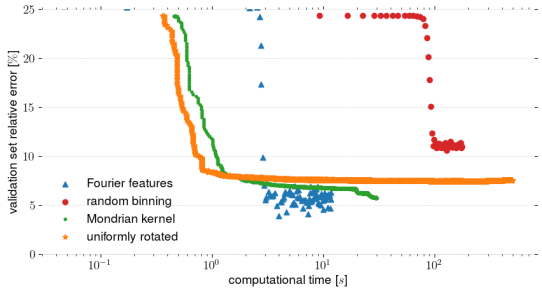
(1) Convergence of Random Feature Kernel to Limiting Kernel for Uniform Points on $[0, 1]^2$.



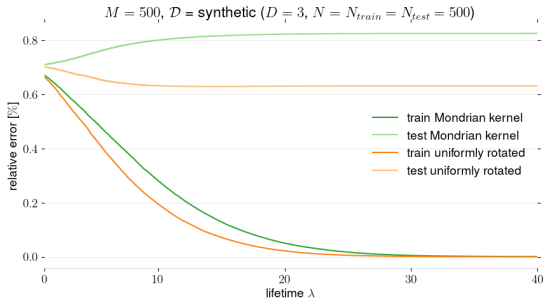
(2) Recovery of Ground Truth Lifetime by Minimizing Validation Error for Synthetic Dataset.



(3) Validation Set Error versus Number of Nonzero Random Features for the `CPU` Dataset.



(4) Validation Set Error versus Computational Time for the `CPU` Dataset.



(5) Relative Error versus Lifetime for the Mondrian Line Dataset.

Figure 4: Five Experiments Evaluating the Uniformly Rotated Mondrian Kernel in Regression Problems.

uniformly rotated Mondrian kernel to benefit from efficient hyperparamter selection as is described in Balog et al. (2016).

In Experiment 3 and Experiment 4, we see that the uniformly rotated Mondrian kernel performs comparably to the Mondrian kernel on the `CPU` data. This is particularly promising given the unique composition of the `CPU` dataset as an *adversarial dataset* for the uniformly rotated Mondrian kernel. The `CPU` dataset has elements of the form $(x_1, \ldots, x_d) \in \mathbb{R}^d$ for $d = 21$, with bounds on each dimension satisfying $0 \leq x_n < 10^4$ for all $n \notin \{3, 8, 9, 20, 21\}$, $0 \leq x_n < 10^5$ for $n \in \{3, 20\}$, and $0 \leq x_n < 10^7$ for $n \in \{8, 9, 21\}$. From this observation, we see that the `CPU` dataset is stretched out across a comparably small number of standard coordinate axes. This suggests that the `CPU` dataset is particularly well-suited for the random feature methods that approximate the Laplace kernel, where the dependence on $||x - x'||_1$ differentiates well between points stretched along the $n \in \{8, 9, 21\}$ axes. Furthermore, this observation is a possible explanation as to why the uniformly rotated Mondrian kernel in Experiment 4 performs slightly worse in the long run on the `CPU` dataset compared to the Mondrian kernel, as the former's isotropic limiting kernel is less effective than the Laplace kernel on this particular dataset.

Furthermore, beyond this difference in limiting kernels, we see that the `CPU` dataset is particularly adversarial for the uniformly rotated Mondrian kernel with respect to *computational intensity*, as generat-

ing a Mondrian tessellation on a particular uniform rotation of the dataset requires, on average, an initial bounding box of higher volume than that of the original dataset. In turn, this requirement demands a higher computational intensity for the uniformly rotated Mondrian kernel on the `CPU` dataset when compared to the Mondrian kernel. With these two factors in mind, it is promising that the uniformly rotated Mondrian kernel has comparable performance to the Mondrian kernel on the `CPU` dataset, a dataset that is well-suited for random feature models with a non-isotropic limiting kernel and, in particular, a dataset that is especially well-suited to the Mondrian kernel.

In Experiment 5, we demonstrate, on a non-adversarial dataset, that the uniformly rotated Mondrian kernel can outperform the Mondrian kernel on the Mondrian line in both the training and testing relative error at all lifetimes $\lambda > 0$. While the `CPU` dataset is stretched out across a comparably small number of standard coordinate axes in $\mathbb{R}^{21}$, the Mondrian line is uniformly distributed across all standard coordinate axes in $\mathbb{R}^3$, making it particularly well-suited for random feature methods that approximate an isotropic kernel. For further experimental evidence that debiasing a model from a particular set of axes leads to improved empirical performance, see Blaser and Fryzlewicz (2016); Ge et al. (2019); Vempala (2012).

In summary, beyond affirming the efficient convergence of the uniformly rotated Mondrian kernel to its limiting kernel and the recovery of a ground truth lifetime, our experiments demonstrate both that the uniformly rotated Mondrian kernel performs comparably to the Mondrian kernel on an adversarial dataset, and that the uniformly rotated Mondrian kernel outperforms the Mondrian kernel on a non-adversarial dataset.

8 CONCLUSION

The uniformly rotated Mondrian kernel as constructed in this work takes advantage of the computationally efficient Mondrian process to compute an isotropic kernel. The uniformly rotated Mondrian kernel converges uniformly to its limiting kernel at an exponential rate comparable to the Mondrian kernel. From our experiments, we observe on an adversarial dataset that the additional computational time spent on rotating the data to construct the uniformly rotated Mondrian kernel does not diminish the high efficiency of the Mondrian process, in addition to observing that the uniformly rotated Mondrian kernel can outperform the Mondrian kernel on a non-adversarial dataset.

One extension of this work is to study an isotropic variant of *Mondrian forests* as in Lakshminarayanan et al. (2014), a variation of the Mondrian kernel that is also based on generating M independent Mondrian tessellations. As in our work above, we anticipate that a *uniformly rotated Mondrian forest* would take advantage of the efficiency of the Mondrian process while approximating an isotropic kernel. For experiments studying Mondrian forests, see Balog et al. (2016).

Finally, we expect that the lifting technique used in our proof of Equation (7) can be used to obtain more general results on the circumradius of the typical cell of Poisson and doubly stochastic Poisson hyperplane processes with discrete directional distributions, aiding the analysis of other transformed Mondrian estimators. Of particular interest is the case where the discrete directional distributions are informed by the training dataset, as such data-driven transformations could further improve the performance of the Mondrian kernel, e.g., the recent work of Baptista et al. (2024) with Mondrian forests.

Acknowledgments

CO is grateful for support from the California Institute of Technology through the Summer Undergraduate Research Fellowship. EO is grateful for support from NSF Grant DMS-2402234.

References

Balog, M., Lakshminarayanan, B., Ghahramani, Z., Roy, D. M., and Teh, Y. W. (2016). The mondrian kernel. In *Proceedings of the 32nd Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 32–41.

Balog, M. and Teh, Y. W. (2015). The Mondrian process for machine learning. *Preprint arXiv:1507.05181*.

Baptista, R., O'Reilly, E., and Xie, Y. (2024). TrIM: Transformed iterative Mondrian forests for gradient-based dimension reduction and high-dimensional regression. *Preprint arXiv:2407.09964*.

Blaser, R. and Fryzlewicz, P. (2016). Random rotation ensembles. *Journal of Machine Learning Research*, 17(1):126–151.

Cattaneo, M. D., Klusowski, J. M., and Underwood, W. G. (2023). Inference with mondrian random forests.

Ge, S., Wang, S., Teh, Y. W., Wang, L., and Elliott, L. (2019). Random tessellation forests. *Advances in Neural Information Processing Systems*, 32:9571–9581.

Hug, D. and Schneider, R. (2007). Asymptotic shapes of large cells in random tessellations. *GAFA Geometric And Functional Analysis*, 17(1):156–191.

Lakshminarayanan, B., Roy, D. M., and Teh, Y. W. (2014). Mondrian forests: Efficient online random forests. In *Advances in neural information processing systems*, pages 3140–3148.

Liu, F., Huang, X., Chen, Y., and Suykens, J. A. K. (2022). Random features for kernel approximation: A survey on algorithms, theory, and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):7128–7148.

Mourtada, J., Gaïffas, S., and Scornet, E. (2020). Minimax optimal rates for mondrian trees and forests. *Annals of Statistics*, 28(4):2253–2276.

Nagel, W. and Weiss, V. (2005). Crack STIT tessellations: Characterization of stationary random tessellations stable with respect to iteration. *Advances in Applied Probability*, 37:859–883.

O'Reilly, E. and Tran, N. M. (2021). Minimax rates for high-dimensional random tessellation forests. *Preprint arXiv:2109.10541*.

O'Reilly, E. and Tran, N. M. (2022). Stochastic geometry to generalize the mondrian process. *SIAM Journal on Mathematics of Data Science*, 4(2):531–552.

Rahimi, A. and Recht, B. (2008). Random features for large-scale kernel machines. *Advances in Neural Information Processing Systems*, 20.

Schneider, R. and Weil, W. (2008). *Stochastic and Integral Geometry*. Springer Berlin Heidelberg.

Schreiber, T. and Thäle, C. (2013). Geometry of iteration stable tessellations: Connection with poisson hyperplanes. *Bernoulli*, 19(5A).

Vempala, S. S. (2012). Randomly-oriented k-d trees adapt to intrinsic dimension. In D'Souza, D., Kavitha, T., and Radhakrishnan, J., editors, *IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS 2012)*, volume 18 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 48–57, Dagstuhl, Germany. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

Wu, L., Yen, I. E., Chen, J., and Yan, R. (2016). Revisiting random binning features: Fast convergence and strong parallelizability. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, page 1265–1274.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
   ⋆ *See Section 7.1 for complete details on our experiments.*
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [No]
   ⋆ *Complexity analysis has not been done in either Balog et al. (2016) or Rahimi and Recht (2008) for random feature models and would require extensive further work.*
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes]
   ⋆ *See Section B and* `README.md` *for source code and dependencies.*

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. [Yes]
   ⋆ *See the statement of Theorem 1, Theorem 2, Proposition 1, and Lemma 1, as well as the preceding paragraphs that explain all technical terminology.*
   (b) Complete proofs of all theoretical results. [Yes]
   ⋆ *See Section 6, Section A.1, Section A.2, and Section A.3 for complete proofs of all theoretical results.*
   (c) Clear explanations of any assumptions. [Yes]
   ⋆ *The statement of all results include their full necessary assumptions.*

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes]
   ⋆ *See Section B and* `README.md`.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes]
   ⋆ *See Section 7.1.*
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes]
   ⋆ *See Section B,* `README.md`*, and Section 7.1.*
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator if your work uses existing assets. [Yes]
   ⋆ *Our figures are all independently created, and the framework of our code from Balog et al. (2016) is cited with its original license in* `NOTICE.md` *and with further detail in* `README.md`.
   (b) The license information of the assets, if applicable. [Yes]
   ⋆ *The license information of the framework of our code from Balog et al. (2016) is included in* `NOTICE.md` *and further detailed in* `README.md`.
   (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
   (d) Information about consent from data providers/curators. [Not Applicable]
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. [Not Applicable]
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

The Uniformly Rotated Mondrian Kernel Supplementary Materials

A APPENDIX

In this Appendix, we will provide detailed proofs of our theoretical results from above.

A.1 Proof of Equation (5)

Proof of Equation (5). Let X be the superposition of M independent rotated Poisson Manhattan tessellations $X_1, \ldots, X_M$ all with intensity $\lambda > 0$. For $n = 1, \ldots, M$, note that the Poisson Manhattan tessellation X_n has directional distribution

$$\varphi_n = \frac{1}{d} \sum_{i=1}^{d} \delta_{u_{n,i}}$$

for some orthonormal set of vectors $\{u_{n,1}, \ldots, u_{n,d}\}$, and thus the associated zonoid of X_n has support function

$$h_{\Pi_{X_n}}(u) = \frac{\lambda}{2} \int_{S^{d-1}} |\langle u, v \rangle| \, \mathrm{d}\varphi_n(v) = \left(\frac{\lambda}{2d} \right) \sum_{i=1}^{d} |\langle u, u_{n,i} \rangle| \, .$$

The above quantity achieves the same maximum and minimum values over the unit sphere $||u||_2 = 1$ as

$$\left(\frac{\lambda}{2d} \right) \sum_{i=1}^{d} |\langle u, e_i \rangle| = \left(\frac{\lambda}{2d} \right) ||u||_1.$$

Thus, by the inequality $||u||_2 \leq ||u||_1 \leq \sqrt{d}||u||_2$ that is tight over the unit sphere $||u||_2 = 1$, we have that

$$\min_{u \in S^{d-1}} h_{\Pi_{X_n}}(u) = \frac{\lambda}{2d}, \quad \max_{u \in S^{d-1}} h_{\Pi_{X_n}}(u) = \frac{\lambda}{2\sqrt{d}}.$$

From Page 158 of Schneider and Weil (2008), we have that the associated zonoid of the superposition X of the independent stationary Poisson hyperplane processes $X_1, \ldots, X_M$ is additive in the sense that

$$\Pi_X = \sum_{n=1}^{M} \Pi_{X_n},$$

where the associated sum is the Minkowski sum of the associated zonoids, i.e., we have that $\Pi_X = \{v_1 + \cdots + v_M : v_n \in \Pi_{X_n}, 1 \leq n \leq M\}$. Furthermore, as the support function of the Minkowski sum of convex bodies is additive, it follows that

$$R(\Pi_X) \leq \max_{u \in S^{d-1}} h_{\Pi_X}(u) = \max_{u \in S^{d-1}} \sum_{n=1}^{M} h_{\Pi_{X_n}}(u) \leq \sum_{n=1}^{M} \max_{u \in S^{d-1}} h_{\Pi_{X_n}}(u) = \frac{\lambda M}{2\sqrt{d}},$$

$$r(\Pi_X) \geq \min_{u \in S^{d-1}} h_{\Pi_X}(u) = \min_{u \in S^{d-1}} \sum_{n=1}^{M} h_{\Pi_{X_n}}(u) \geq \sum_{n=1}^{M} \min_{u \in S^{d-1}} h_{\Pi_{X_n}}(u) = \frac{\lambda M}{2d}.$$

From Theorem 10.3.3 and (10.4) in Schneider and Weil (2008), it holds that $\mathbb{E}[V(Z)] = V(\Pi_X)^{-1}$.

Thus, we have that

$$\mathbb{E}\left[V(Z)\right] \leq \left(\kappa_d \cdot r(\Pi_X)^d\right)^{-1} \leq \frac{1}{\kappa_d}\left(\frac{2d}{\lambda M}\right)^d,$$
$$\mathbb{E}\left[V(Z)\right] \geq \left(\kappa_d \cdot R(\Pi_X)^d\right)^{-1} \geq \frac{1}{\kappa_d}\left(\frac{2\sqrt{d}}{\lambda M}\right)^d.$$

This completes our computation of bounds for the expected volume of the typical cell of the superposition of independent rotated Mondrian tessellations. □

A.2 Proof of Equation (7)

To prove Equation (7), we will first provide an explicit lift for the superposition of independent rotated Mondrian tessellations. While similar to Theorem 3.1 in O'Reilly and Tran (2022) in providing a lift for a STIT tessellation with finitely many cut directions, our result below gives both an explicit characterization of the subspace U and the Poisson Manhattan tessellation $\widetilde{X}$ in the lifted space for our specific setting.

Proposition 1. *Let X be the superposition of M independent rotated Poisson Manhattan tessellations $X_1, \ldots, X_M$ all with intensity $\lambda > 0$ in $\mathbb{R}^d$, where each X_n has directional distribution*

$$\varphi_n = \frac{1}{d}\sum_{i=1}^{d} \delta_{u_{n,i}}$$

for an orthonormal set of vectors $\{u_{n,1}, \ldots, u_{n,d}\}$. It holds that X is the same in distribution as $\widetilde{X} \cap U$, where $\widetilde{X}$ is an axis-aligned Poisson Manhattan tessellation in $\mathbb{R}^{Md}$ with intensity $\widetilde{\lambda} = M^{3/2}\lambda > 0$ and $U = Im(T)$ is a d-dimensional linear subspace of $\mathbb{R}^{Md}$, where the orthogonal transformation $T : \mathbb{R}^d \to \mathbb{R}^{Md}$ is represented by the matrix

$$\mathcal{M}(T) = \frac{1}{\sqrt{M}}\begin{bmatrix} | & & | & & | & & | \\ u_{1,1} & \ldots & u_{1,d} & \ldots & u_{M,1} & \ldots & u_{M,d} \\ | & & | & & | & & | \end{bmatrix}^T. \tag{8}$$

Proof of Proposition 1: First, we recall from Page 158 of Schneider and Weil (2008) that the directional distribution of X is given by

$$\varphi = \frac{1}{M}\sum_{n=1}^{M}\varphi_n = \frac{1}{Md}\sum_{n=1}^{M}\sum_{i=1}^{d}\delta_{u_{n,i}}.$$

We then have from (4.60) in Schneider and Weil (2008) that the intensity of the point process $X \cap \mathrm{Span}(u)$ for any $u \in S^{d-1}$ is given by twice the support function $h_{\Pi_X}(u)$ of the associated zonoid Π_X of X, i.e., we have that

$$\lambda_{X\cap \mathrm{Span}(v)} = M\lambda\int_{S^{d-1}} |\langle u, v\rangle| \mathrm{d}\varphi(v) = \frac{\lambda}{d}\sum_{n=1}^{M}\sum_{i=1}^{d}|\langle u, u_{n,i}\rangle|$$

as in our computation in proving Equation (5). Similarly, for any $\widetilde{u} \in S^{Md-1}$, recalling that

$$\widetilde{\varphi} = \frac{1}{Md}\sum_{k=1}^{Md}\delta_{e_k}$$

is the directional distribution of $\widetilde{X}$, we have that

$$\lambda_{\widetilde{X}\cap \mathrm{Span}(\widetilde{u})} = M^{3/2}\lambda\int_{S^{Md-1}} |\langle \widetilde{u}, v\rangle| \mathrm{d}\widetilde{\varphi}(v) = \frac{\lambda\sqrt{M}}{d}\sum_{k=1}^{Md}|\langle \widetilde{u}, e_k\rangle| = \frac{\lambda\sqrt{M}}{d}||\widetilde{u}||_1.$$

We will then consider the linear transformation $T : \mathbb{R}^d \to \mathbb{R}^{Md}$ as in (8). We note for any $u \in \mathbb{R}^d$ that

$$Tu = \frac{1}{\sqrt{M}}\left(\langle u, u_{1,1}\rangle, \ldots, \langle u, u_{1,d}\rangle, \ldots, \langle u, u_{M,1}\rangle, \ldots, \langle u, u_{M,d}\rangle\right),$$

and so we have that

$$\lambda_{X\cap \mathrm{Span}(u)} = \frac{\lambda}{d}\sum_{i=1}^{M}\sum_{n=1}^{d}|\langle u, u_{n,i}\rangle| = \frac{\lambda\sqrt{M}}{d}\,||Tu||_1 = \lambda_{\widetilde{X}\cap \mathrm{Span}(Tu)}.$$

By uniqueness on Page 131 of Schneider and Weil (2008), it follows that X is distributed the same as $\widetilde{X}\cap U$, where $U = \mathrm{Im}(T)$. Finally, we note since each $\{u_{n,1},\ldots,u_{n,d}\}$ is an orthonormal set of vectors, the matrix

$$\begin{bmatrix} | & \cdots & | \\ u_{n,1} & \cdots & u_{n,d} \\ | & \cdots & | \end{bmatrix}^T = \begin{bmatrix} — & u_{n,1} & — \\ \vdots & \vdots & \vdots \\ — & u_{n,d} & — \end{bmatrix}$$

is orthogonal and

$$\left\langle (u_{n,1}^{(j_1)},\ldots,u_{n,d}^{(j_1)}), (u_{n,1}^{(j_2)},\ldots,u_{n,d}^{(j_2)})\right\rangle = \begin{cases} 1, & j_1 = j_2, \\ 0, & \text{otherwise}, \end{cases}$$

where $u_{n,i} = (u_{n,i}^{(1)},\ldots,u_{n,i}^{(d)})$ are the coordinates of the vector $u_{n,i}$. We thus have for all $1\le j_1, j_2 \le d$ that

$$\begin{aligned} &\langle Te_{j_1}, Te_{j_2}\rangle \\ &= \frac{1}{M}\left(u_{1,1}^{(j_1)}u_{1,1}^{(j_2)} + \cdots + u_{1,d}^{(j_1)}u_{1,d}^{(j_2)} + \cdots + u_{M,1}^{(j_1)}u_{M,1}^{(j_2)} + \cdots + u_{M,d}^{(j_1)}u_{M,d}^{(j_2)}\right) \\ &= \frac{1}{M}\left(\left\langle (u_{1,1}^{(j_1)},\ldots,u_{1,d}^{(j_1)}), (u_{1,1}^{(j_2)},\ldots,u_{1,d}^{(j_2)})\right\rangle + \cdots + \left\langle (u_{M,1}^{(j_1)},\ldots,u_{M,d}^{(j_1)}), (u_{M,1}^{(j_2)},\ldots,u_{M,d}^{(j_2)})\right\rangle\right) \\ &= \begin{cases} 1, & j_1 = j_2 \\ 0, & \text{otherwise}. \end{cases} \end{aligned}$$

We thus have that $T:\mathbb{R}^d\to\mathbb{R}^{Md}$ is an orthogonal transformation, and so U is a d-dimensional linear subspace of $\mathbb{R}^{Md}$ such that T preserves the structure of X as a random tessellation as is desired. □

Proof of Equation (7). We will let $\widetilde{X}$ and $U = \mathrm{Im}(T)$ be as defined in Proposition 1 for X a superposition of M independent rotated Poisson Manhattan tessellations $X_1,\ldots,X_M$ so that X and $\widetilde{X}\cap U$ are the same in distribution. We have in particular for the typical cell Z of X that

$$\mathbb{P}[R(Z)\ge a] = \mathbb{P}\left[R\left(U\cap\widetilde{Z}\right)\ge a\right],$$

where the typical cell $\widetilde{Z}$ of $\widetilde{X}$ is distributed as

$$\prod_{k=1}^{Md}[-t_k, t_k].$$

We recall that the particular orthogonal lift $T:\mathbb{R}^d\to\mathbb{R}^{Md}$ of X is represented by the matrix

$$\mathcal{M}(T) = \left[u_1,\ldots,u_d\right] = \frac{1}{\sqrt{M}}\begin{bmatrix} | & & | & & | & & | \\ u_{1,1} & \cdots & u_{1,d} & \cdots & u_{M,1} & \cdots & u_{M,d} \\ | & & | & & | & & | \end{bmatrix}^T$$

for orthonormal sets of vectors $\{u_{1,i},\ldots,u_{d,i}\}$ with each $u_{n,i}\in\mathbb{R}^d$ for all $1\le i\le M$. Furthermore, we have for all fixed $t_1,\ldots,t_{Md}\in(0,\infty)$ that

$$R\left(U\cap\prod_{k=1}^{Md}[-t_k,t_k]\right) = \max\left\{||x||_2 : x\in U\cap\prod_{k=1}^{Md}[-t_k,t_k]\right\}$$

$$= \max\left\{ ||c||_2 : \sum_{i=1}^{d} c_i u_i \in \prod_{k=1}^{Md} [-t_k, t_k] \right\},$$

where we recall since T is orthogonal that $u_1, \dots, u_d \in \mathbb{R}^{Md}$ is orthogonal. We will then define

$$\begin{bmatrix} | & & | \\ u_{n,1} & \dots & u_{n,d} \\ | & & | \end{bmatrix}^T = \begin{bmatrix} | & & | \\ v_{n,1} & \dots & v_{n,d} \\ | & & | \end{bmatrix}$$

so that

$$\mathcal{M}(T) = \frac{1}{\sqrt{M}} \begin{bmatrix} | & & | & & | & & | \\ u_{1,1} & \dots & u_{1,d} & \dots & u_{M,1} & \dots & u_{M,d} \\ | & & | & & | & & | \end{bmatrix}^T = \frac{1}{\sqrt{M}} \begin{bmatrix} | & & | \\ v_{1,1} & \dots & v_{1,d} \\ | & & | \\ \vdots & \vdots & \vdots \\ | & & | \\ v_{M,1} & \dots & v_{M,d} \\ | & & | \end{bmatrix}$$

$$= \begin{bmatrix} u_1 & \dots & u_d \end{bmatrix},$$

where each $v_{n,1}, \dots, v_{n,d} \in \mathbb{R}^d$ is orthonormal. It follows for any

$$\sum_{i=1}^{d} c_i u_i \in \prod_{k=1}^{Md} [-t_k, t_k],$$

that

$$\sum_{i=1}^{d} c_i u_i^{(k)} \in [-t_k, t_k] \text{ for all } 1 \le k \le Md$$

$$\Rightarrow \sum_{i=1}^{d} c_i \left(\frac{v_{n,i}}{\sqrt{M}} \right) \in \prod_{k=nd+1}^{(n+1)d} [-t_k, t_k] \text{ for all } 1 \le n \le M$$

$$\Rightarrow \left\| \sum_{i=1}^{d} c_i \left(\frac{v_{n,i}}{\sqrt{M}} \right) \right\|_2 \overset{(\dagger)}{=} \frac{1}{\sqrt{M}} ||c||_2 \le \sqrt{t_{nd+1}^2 + \dots + t_{(n+1)d}^2} \text{ for all } 1 \le n \le M,$$

where $(\dagger)$ holds since each set $\{v_{n,1}, \dots, v_{n,d}\}$ is orthogonal. We have thus computed an upper bound of the form

$$R\left(U \cap \prod_{k=1}^{Md} [-t_k, t_k] \right) = \max\left\{ ||c||_2 : \sum_{n=1}^{d} c_n u_n \in \prod_{n=1}^{Md} [-t_k, t_k] \right\}$$

$$\le \min\left(\sqrt{t_1^2 + \dots + t_d^2}, \dots, \sqrt{t_{(M-1)d+1}^2 + \dots + t_{Md}^2} \right) \sqrt{M}$$

$$\le \min\left(t_1 + \dots + t_d, \dots, t_{(M-1)d+1} + \dots + t_{Md} \right) \sqrt{M}$$

for any particular set of rotated Poisson Manhattan tessellations $X_1, \dots, X_M$. Finally, since $\widetilde{X}$ is a Poisson Manhattan tessellation in $\mathbb{R}^{Md}$ of intensity $\lambda M^{3/2} > 0$, we know that $t_k \sim \text{Exp}(2\lambda\sqrt{M}/d)$. It follows that each $t_k\sqrt{M} \sim \text{Exp}(2\lambda/d)$ so that

$$(t_{id+1} + \dots + t_{(i+1)d})\sqrt{M} \sim \text{Erlang}\left(d, \frac{2\lambda}{d} \right),$$

$$\text{where } \mathbb{P}\left[\text{Erlang}\left(d, \frac{2\lambda}{d}\right) \geq a\right] = \exp\left(-\frac{2\lambda a}{d}\right) \sum_{n=0}^{d-1} \frac{1}{n!} \left(\frac{2\lambda a}{d}\right)^n.$$

Finally, since for all independent and identically distributed random variables $Y_1, \ldots, Y_n$ it holds that

$$\mathbb{P}[\min(Y_1, \ldots, Y_n) \geq a] = \mathbb{P}[Y_1 \geq a]^n,$$

we have that

$$\mathbb{P}\left[R(Z) \geq a\right] \leq \exp\left(-\frac{2M\lambda a}{d}\right) \left(\sum_{n=0}^{d-1} \frac{1}{n!} \left[\frac{2\lambda a}{d}\right]^n\right)^M$$

as is desired, completing our computation of an upper bound for the circumradius of the typical cell of the superposition of independent rotated Mondrian tessellations. □

A.3 Proof of Theorem 2

To prove Theorem 2, we will need to make use of the application of Campbell's theorem that was stated above.

Theorem 3. (Schneider and Weil, 2008, Equation 4.3) *For any nonnegative measurable function f on the space of nonempty compact sets in $\mathbb{R}^d$ and for any stationary random tessellation X, we have that*

$$\mathbb{E}\left[\sum_{C \in cells(X)} f(C)\right] = \frac{1}{\mathbb{E}[V(Z)]} \cdot \mathbb{E}\left[\int_{\mathbb{R}^d} f(Z + y) \mathrm{d}y\right],$$

where Z is the typical cell of the X.

Proof of Theorem 2. We will consider the definitions and events as are detailed in the outline from Section 6. From $A_1 \cap A_2 \cap A_3$, we have for every $x \in \mathcal{X}$ that there is a point $u \in \mathcal{U}$ such that x, u are in the same cell of X and

$$||x - u||_2 \leq \frac{\delta}{4\lambda\sqrt{d}}.$$

Then, for any $x, x' \in \mathcal{X}$ we have that

$$|k_M(x, x') - k_\infty(x, x')| \leq |k_M(x, x') - k_M(u, u')| + |k_M(u, u') - k_\infty(u, u')| + |k_\infty(u, u') - k_\infty(x, x')|,$$

where $u, u' \in \mathcal{U}$ are as described above. Since x, u and x', u' are in the same cells of X,

$$k_M(x, x') = k_M(u, u') \Rightarrow |k_M(x, x') - k_M(u, u')| = 0$$

for all $x, x' \in \mathcal{X}$. We then note from A_4 that $\sup_{x,x' \in \mathcal{X}} |k_M(u, u') - k_\infty(u, u')| \leq \delta/2$. We finally note from Theorem 1 that

$$\begin{aligned}
\sup_{x,x' \in \mathcal{X}} |k_\infty(u, u') - k_\infty(x, x')| &= \sup_{x,x' \in \mathcal{X}} \frac{1}{\omega_d} \left| \int_{S^{d-1}} \left(e^{-\lambda ||u-u'||_2 ||v||_1} - e^{-\lambda ||x-x'||_2 ||v||_1}\right) \mathrm{d}v \right| \\
&\leq \sup_{x,x' \in \mathcal{X}} \frac{1}{\omega_d} \int_{S^{d-1}} \left| e^{-\lambda ||u-u'||_2 ||v||_1} - e^{-\lambda ||x-x'||_2 ||v||_1} \right| \mathrm{d}v \\
&\overset{(a)}{\leq} \sup_{x,x' \in \mathcal{X}} \lambda \left| ||u - u'||_2 - ||x - x'||_2 \right| \frac{1}{\omega_d} \int_{S^{d-1}} ||v||_1 \, \mathrm{d}v \\
&\overset{(b)}{\leq} \sup_{x,x' \in \mathcal{X}} \lambda\sqrt{d} \left| ||u - u'||_2 - ||x - x'||_2 \right| \frac{1}{\omega_d} \int_{S^{d-1}} ||v||_2 \, \mathrm{d}v \\
&\overset{(c)}{=} \sup_{x,x' \in \mathcal{X}} \lambda\sqrt{d} \left| ||u - u'||_2 - ||x - x'||_2 \right|
\end{aligned}$$

$$\leq \sup_{x,x' \in \mathcal{X}} \lambda\sqrt{d}\, ||(u-u') - (x-x')||_2$$

$$\leq \sup_{x,x' \in \mathcal{X}} \lambda\sqrt{d}\, (||u-x||_2 + ||u'-x'||_2) \leq \frac{\delta}{2},$$

where (a) follows from the inequality $\left|e^{-|x|} - e^{-|y|}\right| \leq |x-y|$, (b) follows from the inequality $||v||_1 \leq \sqrt{d}\, ||v||_2$, and (c) holds from

$$\frac{1}{\omega_d} \int_{S^{d-1}} ||v||_2 \, \mathrm{d}v = \frac{1}{\omega_d} \int_{S^{d-1}} \mathrm{d}v = 1.$$

We thus have that the δ-approximation on $\mathcal{X}$ is satisfied on the event $A_1 \cap A_2 \cap A_3 \cap A_4$, and so

$$\mathbb{P}\left[\sup_{x,x' \in \mathcal{X}} |k_M(x,x') - k_\infty(x,x')| > \delta\right] \leq \mathbb{P}(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c)$$

$$\leq \mathbb{P}(A_1^c) + \mathbb{P}(A_2^c) + \mathbb{P}(A_3^c) + \mathbb{P}(A_4^c),$$

from a union bound. We will now bound these four probabilities. First, we saw from the outline above that

$$\mathbb{P}(A_1^c) \leq \left(\frac{\kappa_d^2 d}{2\lambda M}\right)\left(\frac{\lambda M}{2\sqrt{d}}\right)^d e^{-2\lambda M(R-r)}.$$

Second, we see that for A_2^c,

$$\mathbb{P}(A_2^c) = \mathbb{P}[\text{there exists a cell in } X \text{ with circumcenter in } B_R \text{ disjoint from } \mathcal{U}]$$

$$= \mathbb{P}\left[\bigcup_{\text{cells } C \in X} (c(C) \in B_R) \cap (C \text{ is disjoint from } \mathcal{U})\right]$$

$$\leq \mathbb{E}\left[\sum_{\text{cells } C \in X} \mathbb{1}[c(C) \in B_R] \mathbb{1}\left[C \text{ is disjoint from } \mathcal{U}\right]\right]$$

$$= \left(\frac{1}{\mathbb{E}[V(Z)]}\right) \mathbb{E}\left[\int_{\mathbb{R}^d} \mathbb{1}[y \in B_R] \mathbb{1}\left[Z + y \text{ is disjoint from } \mathcal{U}\right] \mathrm{d}y\right],$$

where we have again used Campbell's Theorem in the last equality. Since $r(C) \geq \varepsilon/2\sqrt{d}$ implies that C contains a point of $\mathcal{U}$, we note that

$$\mathbb{P}(A_2^c) \leq \left(\frac{1}{\mathbb{E}[V(Z)]}\right) \mathbb{E}\left[\int_{\mathbb{R}^d} \mathbb{1}[y \in B_R] \mathbb{1}[r(Z+y) < \varepsilon/2\sqrt{d}] \mathrm{d}y\right] = \left(\frac{\kappa_d R^d}{\mathbb{E}[V(Z)]}\right) \mathbb{P}[r(Z) < \varepsilon/2\sqrt{d}]$$

$$\leq \kappa_d^2 \left(\frac{\lambda R M}{2\sqrt{d}}\right)^d \mathbb{P}[r(Z) < \varepsilon/2\sqrt{d}]$$

$$= \kappa_d^2 \left(\frac{\lambda R M}{2\sqrt{d}}\right)^d \left(1 - e^{-M\lambda\varepsilon\sqrt{d}}\right)$$

$$\leq \kappa_d^2 \left(\frac{\lambda R M}{2\sqrt{d}}\right)^d M\lambda\varepsilon\sqrt{d},$$

where in the last inequality we have used the fact that $1 - e^{-x} \leq x$ for all $x \in \mathbb{R}$.

Third, we see that for A_3^c,

$$\mathbb{P}(A_3^c) = \mathbb{P}[\text{there exists a cell in } X \text{ with circumcenter in } B_R \text{ and diameter } \geq \delta/4\lambda\sqrt{d}]$$

$$= \mathbb{P}\left[\bigcup_{\text{cells } C\in X} (c(C) \in B_R) \cap \left(R(C) \geq \frac{\delta}{8\lambda\sqrt{d}}\right)\right]$$

$$\leq \sum_{\text{cells } C\in X} \mathbb{P}\left[c(C) \in B_R\right] \mathbb{P}\left[R(C) \geq \frac{\delta}{8\lambda\sqrt{d}}\right]$$

$$= \mathbb{E}\left[\sum_{\text{cells } C\in X} \mathbb{1}\left[c(C) \in B_R\right] \mathbb{1}\left[R(C) \geq \frac{\delta}{8\lambda\sqrt{d}}\right]\right]$$

$$= \left(\frac{1}{\mathbb{E}\left[V(Z)\right]}\right) \mathbb{E}\left[\int_{\mathbb{R}^d} \mathbb{1}\left[y \in B_R\right] \mathbb{1}\left[R(Z+y) \geq \frac{\delta}{8\lambda\sqrt{d}}\right] \mathrm{d}y\right]$$

$$= \left(\frac{\kappa_d R^d}{\mathbb{E}\left[V(Z)\right]}\right) \mathbb{P}\left[R(Z) \geq \frac{\delta}{8\lambda\sqrt{d}}\right],$$

with $R(C+y) = R(C)$ for all cells C since the circumradius is translation invariant. We thus have from (7) that

$$\mathbb{P}(A_3^c) \leq \kappa_d^2 \left(\frac{\lambda R M}{2\sqrt{d}}\right)^d \left(\sum_{n=0}^{d-1} \frac{1}{n!} \left[\frac{\delta}{4d^{3/2}}\right]^n\right)^M e^{-M\delta/4d^{3/2}}.$$

Finally, we see that for A_4^c by Hoeffding's inequality that

$$\mathbb{P}(A_4^c) = \mathbb{P}[\text{the } \delta/2\text{-approximation fails on } \mathcal{U}] = \mathbb{P}\left[\text{there exists } u, u' \in \mathcal{U} \text{ such that } |k_M(u,u') - k_\infty(u,u')| > \delta/2\right]$$

$$= \mathbb{P}\left[\bigcup_{u,u'\in\mathcal{U}} |k_M(u,u') - k_\infty(u,u')| > \delta/2\right]$$

$$\leq \sum_{u,u'\in\mathcal{U}} \mathbb{P}\left[|k_M(u,u') - k_\infty(u,u')| > \delta/2\right] \leq \sum_{u,u'\in\mathcal{U}} 2\exp\left(-M\delta^2/2\right).$$

For any ε-grid covering $\mathcal{U}$ of $B_{R+\varepsilon}$, we have that

$$\#(\mathcal{U}) \leq \prod_{n=1}^{d} \left\lceil \frac{2[R+\varepsilon]}{\varepsilon} \right\rceil \leq \prod_{n=1}^{d} \left(\frac{4[R+\varepsilon]}{\varepsilon}\right) = \left(\frac{4R}{\varepsilon} + 1\right)^d$$

since $\varepsilon < 2r$. We thus have that

$$\mathbb{P}(A_4^c) \leq \left[\left(\frac{4R}{\varepsilon} + 1\right)^d\right]^2 \left(2e^{-M\delta^2/2}\right) = \left(2e^{-M\delta^2/2}\right) \sum_{n=0}^{2d} \binom{2d}{n} \left(\frac{4R}{\varepsilon}\right)^n < \left(2e^{-M\delta^2/2}\right) \left(\sum_{n=0}^{2d} \binom{2d}{n} (4R)^n\right) \varepsilon^{-2d}$$

for all small enough $0 < \varepsilon < 1$. This gives us that

$$\mathbb{P}\left[\sup_{x,x'\in\mathcal{X}} |k_M(x,x') - k_\infty(x,x')| > \delta\right] < \left(\frac{\kappa_d^2 d}{2\lambda M}\right) \left(\frac{\lambda M}{2\sqrt{d}}\right)^d e^{-2\lambda M(R-r)}$$

$$+ \kappa_d^2 \left(\frac{\lambda R M}{2\sqrt{d}}\right)^d M\lambda\varepsilon\sqrt{d}$$

$$+ \kappa_d^2 \left(\frac{\lambda R M}{2\sqrt{d}}\right)^d \left(\sum_{n=0}^{d-1} \frac{1}{n!} \left[\frac{\delta}{4d^{3/2}}\right]^n\right)^M e^{-M\delta/4d^{3/2}}$$

$$+ 2\varepsilon^{-2d}\left(\sum_{n=0}^{2d}\binom{2d}{n}(4R)^n\right)e^{-M\delta^2/2}.$$

The above expression is minimized at

$$\varepsilon_0 = \left(2\kappa_d^{-2}Re^{-M\delta^2/2}\left(\frac{2\sqrt{d}}{\lambda RM}\right)^{d+1}\sum_{n=0}^{2d}\binom{2d}{n}(4R)^n\right)^{1/(2d+1)},$$

and so we have that

$$\begin{aligned}\mathbb{P}\left[\sup_{x,x'\in\mathcal{X}}|k_M(x,x') - k_\infty(x,x')| > \delta\right] &< c_1M^{d-1}e^{-2\lambda M(R-r)} + c_2M^{d+d/(2d+1)}e^{-M\delta^2/(4d+2)} \\ &\quad + c_3M^d e^{M(\alpha(\delta)-\delta/4d^{3/2})} + c_4M^{d+d/(2d+1)}e^{-M\delta^2/(4d+2)}\end{aligned}$$

for constants c_1, c_2, c_3, c_4 not depending on M, δ and function

$$\alpha(\delta) = \log\left(\sum_{n=0}^{d-1}\frac{1}{n!}\left[\frac{\delta}{4d^{3/2}}\right]^n\right)$$

which satisfies $\alpha(\delta) - \delta/4d^{3/2} < 0$ and $(\alpha(\delta) - \delta/4d^{3/2}) \to 0$ slower than $-M\delta^2/(4d+2) \to 0$ as $\delta \to 0$. Thus, letting $R - r > 0$ be sufficiently small, it holds for small enough $\delta > 0$ that

$$\mathbb{P}\left[\sup_{x,x'\in\mathcal{X}}|k_M(x,x') - k_\infty(x,x')| > \delta\right] \in \mathcal{O}\left(M^{d+d/(2d+1)}e^{-M\delta^2/(4d+2)}\right),$$

completing our proof of the rate of uniform convergence of the uniformly rotated Mondrian kernel to its limiting kernel. □

B EXPERIMENTS CODE

The following link is to a GitHub repository with the code to replicate our experiments. The `README.md` file contains full information about the required packages and their respective versions at the time of figure generation.