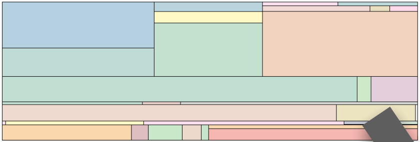
# Random Tessellation Forests

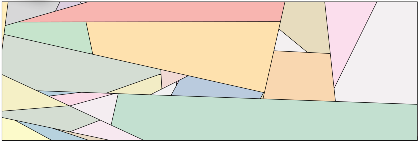
Shufei Ge[1], Shijia Wang[1,2], Yee Whye Teh[3],
Liangliang Wang[1] and Lloyd T. Elliott[1]

[1]Simon Fraser University, Canada
[2]Nankai University, China
[3]University of Oxford, UK

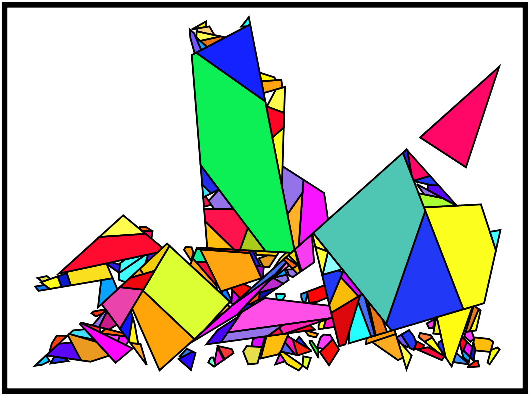


Mondrian process

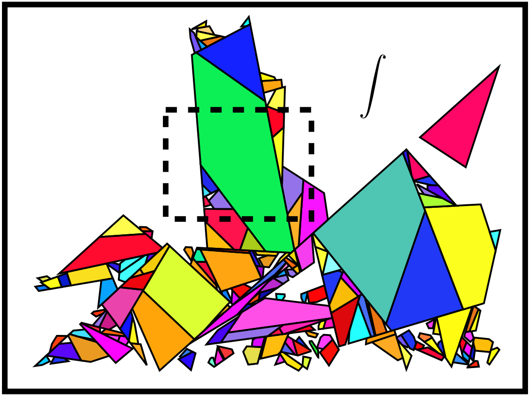
Tessellation process



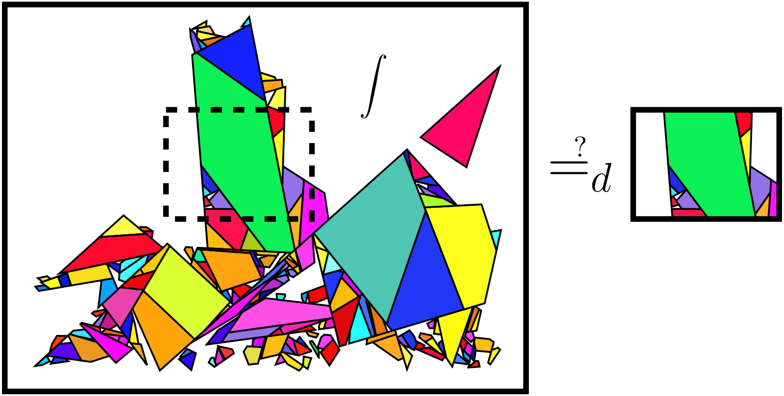
# Projectivity (definition by picture)



# Projectivity (definition by picture)

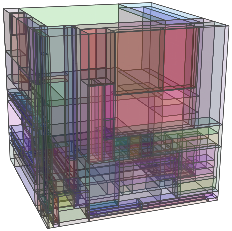


# Projectivity (definition by picture)

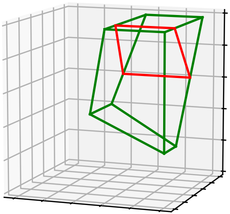


# Related work

- Ostomachion process (Fan et al. 2016)
- Binary space partitioning tree process (Fan et al. 2018)
- Binary space partitioning forests (Fan et al. 2019)
- Stable iterated tessellations (Nagel and Weiss. 2005)
- Mondrian forests (Lakshminarayanan et al. 2014)



Roy and Teh. 2008



Fan et al. 2019

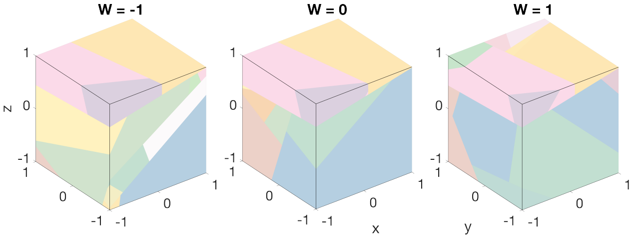
# Notation

- A tessellation $Y(W)$ of a set $W \subseteq \mathbb{R}^d$ is a finite set of polytopes *s.t.*:

$$\bigcup_{a \in Y(W)} a = W, \text{ and } \forall\, a, b \in Y(W), \text{interior}(a) \bigcap \text{interior}(b) = \varnothing.$$

- A polytope is a bounded, nonempty intersection of closed half-planes.

- Let $[S]$ be the set of affine hyperplanes in $\mathbb{R}^d$ intersecting $S \subseteq W$.



# Stable iterated tessellations

Any measure $\Lambda$ on $H = [W]$ induces a tessellation-valued MJP:

**Algorithm 1** Generative Process for RTPs

```
Inputs: a) Bounded domain W, b) RTP measure Λ on H, c) prespecified budget τ.
Outputs: A realisation of the Random Tessellation Process (Y_t)_{0≤t≤τ}.
τ_0 ← 0.
Y_0 ← {W}.
while τ_0 ≤ τ do
    Sample τ' ~ Exp( Σ_{a∈Y_{τ_0}} Λ([a]) ).
    Set Y_t ← Y_{τ_0} for all t ∈ (τ_0, min{τ, τ_0 + τ'}].
    Set τ_0 ← τ_0 + τ'.
    if τ_0 ≤ τ then
        Sample a polytope a from the set Y_{τ_0} with probability proportional to (w.p.p.t.) Λ([a]).
        Sample a hyperplane h from [a] according to the probability measure Λ(· ∩ [a])/Λ([a]).
        Y_{τ_0} ← (Y_{τ_0}/{a}) ∪ {a ∩ h^-, a ∩ h^+}. (h^- and h^+ are the h-bounded closed half planes.)
    else
        return the tessellation-valued right-continuous MJP sample (Y_t)_{0≤t≤τ}.
```

# Conditions for projectivity

- **Theorem [Nagel and Weiss 2005].** If $\Lambda$ is translation invariant and symmetric and supported on an orthogonal set of $d$ hyperplanes, then for all measurable subsets $W' \subseteq W$, $Y(W') =_d Y(W) \cap W'$.

# Random Tessellation Processes

- Every hyperplane $h \in H$ can be written uniquely as:

$$h = \{P : \langle n, P - un \rangle = 0\} \ \ s.t.\ n \in S^{d-1}, u \in \mathbb{R}_{\geq 0}.$$

- Then, $\varphi : S^{d-1} \times \mathbb{R}_{\geq 0} \longmapsto H$ by $\varphi(n, u) = h$ is a bijection.

- Any measure $\Lambda$ on $H$ is induced by a measure $\Lambda \circ \varphi$ on $S^{d-1} \times \mathbb{R}_{\geq 0}$.

# Random Tessellation Processes

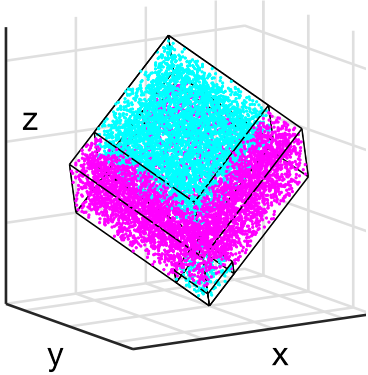
- Let $\Lambda \circ \varphi$ be the product measure $\lambda^d \times \lambda_+$ such that $\lambda^d$ is symmetric and $\lambda_+$ is the Lebesgue measure on $\mathbb{R}_{\geq 0}$.
- **Theorem.** $\Lambda \circ \varphi$ is translation invariant and symmetric. (Proof in the *Supplementary Material*.)
- We refer to such $\Lambda$ as Random Tessellation Process (RTP) measures.
- All RTP measures induce projective tessellations.

# Relation to cutting Bayesian nonparametrics

- The Mondrian process is an RTP with $\lambda^d$ a set of delta functions on the poles of $S^{d-1}$ (MRTP).
- The binary space partitioning tree process is an RTP with $\lambda^d$ the uniform measure on the sphere (uRTP).
- The binary space partitioning forest is an RTP with $\lambda^d$ a convolution between uniform measures and delta functions.
- Weighted versions of these RTPs encode priors over variable importance: *wMRTP*, *wuRTP*.

# Modelling data with RTPs

- For categorical data, we associate beta/Bernoulli parameters to each polytope, yielding an RTP posterior.




The Mondrian cube dataset

# Inference

- We derive a sequential Monte Carlo (SMC) algorithm for inference.
- We consider random forest versions of RTPs (uRTF, MRTF and wuRTF, wMRTF).
- We also implement an efficient RTF in a similar way to the Mondrian forest (Lakshminarayanan et al. 2015), in which likelihoods are dropped from the SMC sampler (uRTF.i and MRTF.i). We also use pausing conditions: $\tau = \infty$, and spherical approximations.

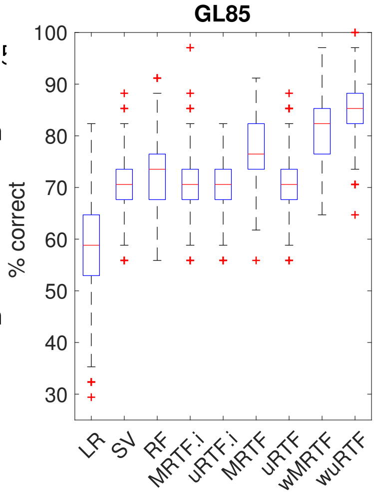
# Inference

**Algorithm 2** SMC for inferring RTP posteriors

**Inputs:** a) Training dataset $\boldsymbol{V}$, $\boldsymbol{Z}$, b) RTP measure $\Lambda$ on $H$, c) prespecified budget $\tau$, d) likelihood hyperparameter $\alpha$.
**Outputs:** Approximate RTP posterior $\sum_{m=1}^{M} \varpi_m \delta_{\triangle_{\tau,m}}$ at time $\tau$. ($\varpi_m$ are particle weights.)
Set $\tau_m \leftarrow 0$, for $m = 1, \dots, M$.
Set $\triangle_{0,m} \leftarrow \{\text{hull } \boldsymbol{V}\}$, $\varpi_m \leftarrow 1/M$, for $m = 1, \dots, M$.
**while** $\min\{\tau_m\}_{m=1}^{M} < \tau$ **do**
  Resample $\triangle'_{\tau_m,m}$ from $\{\triangle_{\tau_m,m}\}_{m=1}^{M}$ *w.p.p.t.* $\{\varpi_m\}_{m=1}^{M}$, for $m = 1, \dots, M$.
  Set $\triangle_{\tau_m,m} \leftarrow \triangle'_{\tau_m,m}$, for $m = 1, \dots, M$.
  Set $\varpi_m \leftarrow 1/M$, for $m = 1, \dots, M$.
  **for** $m \in \{m : m = 1, \dots, M \text{ and } \tau_m < \tau\}$ **do**
    Sample $\tau' \sim \text{Exp}\left(\sum_{a \in \triangle_{\tau_m,m}} r_a\right)$. ($r_a$ is the radius of the smallest closed ball containing $a$.)
    Set $\triangle_{t,m} \leftarrow \triangle_{\tau_m,m}$, for all $t \in (\tau_m, \min\{\tau, \tau_m + \tau'\}]$.
    **if** $\tau_m + \tau' \leq \tau$ **then**
      Sample $a$ from the set $\triangle_{\tau_m,m}$ *w.p.p.t.* $r_a$.
      Sample $h$ from $[a]$ according to $\Lambda(\cdot \cap [a])/\Lambda([a])$ using *Section* 2.2.1.
      Set $\triangle_{\tau_m,m} \leftarrow (\triangle_{\tau_m,m}/\{a\}) \cup \{\text{hull}(\boldsymbol{V} \cap a \cap h^-), \text{hull}(\boldsymbol{V} \cap a \cap h^+)\}$.
      Set $\varpi_m \leftarrow \varpi_m P(\boldsymbol{Z} | \triangle_{\tau_m,m}, \boldsymbol{V}, \alpha)/P(\boldsymbol{Z} | \triangle'_{\tau_m,m}, \boldsymbol{V}, \alpha)$ according to (3).
    **else**
      Set $\triangle_{t,m} \leftarrow \triangle_{\tau_m,m}$, for $t \in (\tau_m, \tau]$.
    Set $\tau_m \leftarrow \tau_m + \tau'$.
  Set $\mathcal{Z} \leftarrow \sum_{m=1}^{M} \varpi_m$.
  Set $\varpi_m \leftarrow \varpi_m / \mathcal{Z}$, for $m = 1, \dots, M$.
**return** the particle approximation $\sum_{m=1}^{M} \varpi_m \delta_{\triangle_{\tau,m}}$.

# Results

- *GL85*: $X$ = gene expression in glioblastoma tissue, $Y$ = astrocytoma grade ($N = P = 8$
- *SCZ42*: $X$ = gene expression in superior temporal cortex, $Y$ = schizophrenia indicator ($N = P = 42$).
- *SCZ51*: $X$ = gene expression in anterior prefrontal cortex, $Y$ = schizophrenia indicator ($N = P = 51$).
- *SCZ93*: *SCZ51*+*SCZ51*.



# Results

| Dataset | BL | LR | SVM | RF | MRTF.i | uRTF.i | MRTF | uRTF | wMRTF | wuRTF |
|---|---|---|---|---|---|---|---|---|---|---|
| *GL85* | 70.34 | 58.13 | 70.34 | 73.01 | 70.74 | 70.06 | 77.09 | 70.60 | 80.57 | **84.90** |
| *SCZ42* | 46.68 | **57.65** | 46.79 | 51.76 | 49.56 | 48.50 | 49.91 | 47.71 | **53.12** | **53.97** |
| *SCZ51* | 46.55 | 51.15 | 46.67 | **57.38** | 52.55 | 48.58 | **57.95** | 44.70 | **58.12** | 49.05 |
| *SCZ93* | 48.95 | **53.05** | 50.15 | 52.45 | 50.23 | 50.24 | 51.80 | 50.34 | **53.12** | **54.99** |

Table: Percent correct. Bold indicates nominal conservative sign test significance.

# Future work

- Relax spherical approximation.
- Hypermanifold cutting.
- Online PG inference.
- Additive regression trees.