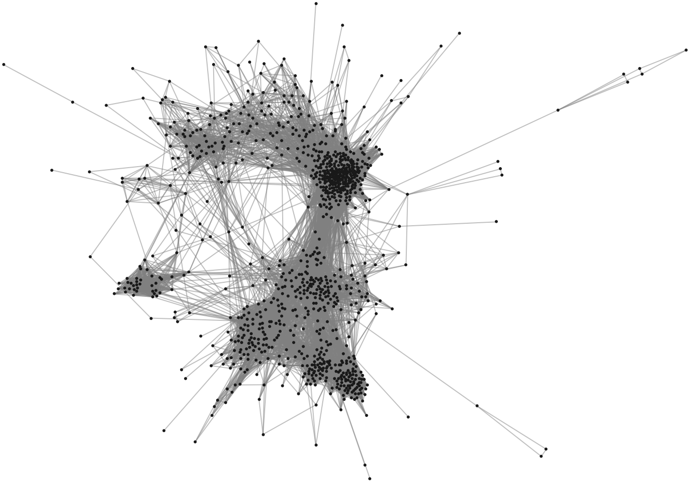
Symmetry and Symmetrization

Peter Orbanz



RECALL THE FUNDAMENTAL LIMIT THEOREMS

Consider an i.i.d. sequence $X = (X_1, X_2, \ldots)$, a function f, and the sample average

$$\frac{1}{n} \sum_{i \leq n} f(X_i) \; .$$

Many forms of estimation/inference/learning involve a sample average somewhere (a plug-in estimate, an empirical risk, a maximum likelihood estimator, etc).

Law of large numbers

$$\text{sample average} \xrightarrow{n \to \infty} \text{expectation of } f \qquad\qquad \text{almost surely}$$

Central limit theorem

$$\sqrt{n}\,(\text{sample average} - \text{expectation of } f) \;\xrightarrow{\mathrm{d}}\; \eta \cdot \text{standard normal}$$

Today's talk in short.

If X is a random structure and its distribution is invariant under a suitable group of transformations, then:

(i) The group tells us how to compute sample averages.

(ii) Versions of the theorems above still hold.

OUTLINE

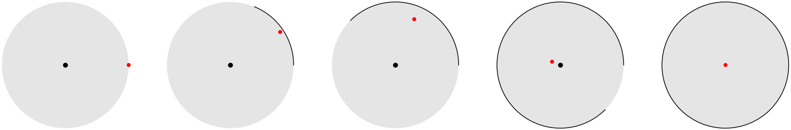
1. Estimation as averaging over transformations.
   Many estimators we use can be written as approximate group averages.

2. The law of large numbers for such estimators.
   Translating results from ergodic theory shows these estimators are consistent.

3. Central limit theorems.
   They satisfy asymptotic normality or other universality results.

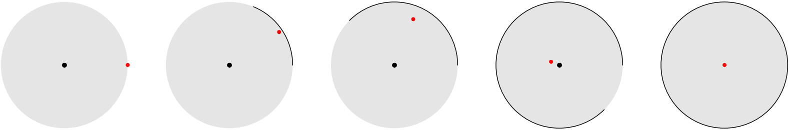
4. de Finetti.
   The estimators converge to expectations under so-called ergodic measures. To explain these, we have to understand de Finetti's theorem and its generalizations.

Averaging Over a Group



x $\quad$ $\mathbf{A}_1 = [0, 3\pi/8]$ $\quad$ $\mathbf{A}_2 = [0, 3\pi/4]$ $\quad$ $\mathbf{A}_3 = [0, 7\pi/4]$ $\quad$ $\mathbf{A}_4 = [0, 2\pi)$

$$x_n = \int_{\mathbf{A}_n} \phi x \, U_n(d\phi) \qquad \text{where } U_n = \text{uniform distribution on } \mathbf{A}_n$$



x $\mathbf{A}_1 = [0, 3\pi/8]$ $\mathbf{A}_2 = [0, 3\pi/4]$ $\mathbf{A}_3 = [0, 7\pi/4]$ $\mathbf{A}_4 = [0, 2\pi)$

Observation I

- If $\mathbf{A}_n = \mathbb{G}$ is the entire group (here: $\mathbf{A}_4$), then x_n is $\mathbb{G}$-invariant.

Observation II

- Suppose a random point X is distributed uniformly on the circle (the boundary of the gray disk).
- Then x_4 is precisely the expected value of X.

If the distribution of X is invariant under the group, then (in this case) $\mathbb{E}[X]$ can be computed by averaging over the group.

- X is infinite i.i.d. random sequence $X = (X_1, X_2, \ldots)$.
- The distribution of X is invariant under permutations or entries.
- f is the function that extracts the first element, $f(X) = X_1$.
- $\mathbf{A}_n$ is the group of permutations of n elements.

$$\int_{\mathbf{A}_n} f(\phi X) U_n(d\phi) \quad = \quad \frac{1}{|\mathbf{A}_n|} \sum_{\phi \in \mathbf{A}_n} X_{\phi(1)} \quad = \quad \frac{1}{n} \sum_{i \leq n} X_i$$

By the law of large numbers, we have

$$\int_{\mathbf{A}_n} f(\phi X) U_n(d\phi) \quad = \quad \frac{1}{n} \sum_{i \leq n} X_i \qquad \xrightarrow{n \to \infty} \qquad \mathbb{E}[X_1]$$

So, at least in this example:

We can compute an expected value of an group-invariant distribution by (approximately) averaging over the group.

First Main Result

Pointwise theorem (Lindenstrauss; Weiss, Tempelman, others)

If the law of X is invariant under a suitable group and $\mathbb{E}[|f(X)|] < \infty$, then

$$\frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n} f(\phi X)|d\phi| \quad \xrightarrow{n\to\infty} \quad \mathbb{E}[f(X)|\,\text{ergodic distribution}]$$

holds almost surely under the law of X.

random structure X	invariance	ergodic distributions	relevant theorem
sequence indexed by $\mathbb{N}$	exchangeability	i.i.d. distributions	de Finetti
graph with vertex set $\mathbb{N}$	exchangeability	graphon distributions	Aldous–Hoover
partition of $\mathbb{N}$	exchangeability	paintbox	Kingman
discrete-time process	stationarity	ergodic process	Rokhlin

The pointwise theorem is a generalized law of large numbers.

On the distribution: Invariance

$$X \stackrel{\mathrm{d}}{=} \phi X \qquad \text{for all } \phi \in \mathbb{G}$$

On the group $\mathbb{G}$: Amenability*

$$\frac{|\mathbf{A}_n \cap \psi \mathbf{A}_n|}{|\mathbf{A}_n|} \xrightarrow{n\to\infty} 1 \qquad \text{for finite/compact sets } \mathbf{A}_n \subset \mathbb{G} \text{ and all } \psi \in \mathbb{G}\ .$$

The most important examples in statistics

group $\mathbb{G}$	invariance	$\mathbf{A}_n$
finite permutations of $\mathbb{N}$	exchangeability	$\mathbb{S}_n$ = permutations of $\{1,\ldots,n\}$
discrete shifts $\mathbb{Z}^d$	stationarity	$\{-n,\ldots,n\}^d$
continuous shifts $\mathbb{R}^d$	stationarity	$[-n,\ldots,n]^d$

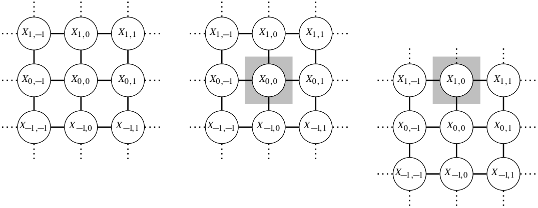
* The sequence $(\mathbf{A}_n)$ must satisfy an extra condition called temperedness (which always holds for a suitable subsequence), and we must substitute ψ by a compact set if $\mathbb{G}$ is not countable, see articles by Lindenstrauss or by Austern/O. in references.

EXAMPLE: STATIONARY FIELD

- X is real-valued random field on the grid $\mathbb{Z}^2$.
- $\mathbb{G}$ is the shift group $\mathbb{Z}^2$.
- $\mathbf{A}_n = \{-n, \ldots, n\}^2$.
- $f(X) = X_{0,0}$

In this case, invariance is stationarity:

$$(X_{i+a,j+b}) \stackrel{\mathrm{d}}{=} (X_{i,j}) \qquad \text{for all indices } (i,j) \in \mathbb{Z}^2 \text{ and shifts } (a,b) \in \mathbb{Z}^2$$



$$\int_{\mathbf{A}_n} f(\phi X) U_n(d\phi) = \frac{1}{|\mathbf{A}_n|} \sum\nolimits_{\phi \in \mathbb{A}_n} f(\phi X) = \frac{1}{(2n+1)^2} \sum\nolimits_{(i,j) \in \{-n,\ldots,n\}^2} X_{ij}$$

GENERAL IDEA

Given:
- X is a random structure that is "very large" (modeled as infinite).
- f is a function that depends only on "a small part" of X.
- A larger (but finite) part of X is observed as a sample.

Task: Estimate $\mathbb{E}[f(X)]$.

$$\text{estimate from sample of size } |\mathbf{A}_n| \;=\; \frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n} f(\phi X)|d\phi|$$

Sets of transformations:

- $\mathbf{A}_n$ = all ϕ needed to cover sample of size n
- $\mathbb{G}$ = all ϕ needed to cover all of X.

The element of $\mathbb{G}$ explain how we collect information from X. The sets $\mathbf{A}_n$ determine what a "sample of size n" is.

Example: Exchangeable Sequence

- X is an infinite random sequence, say in $\mathbb{R}$.
- $\mathbb{G} = \mathbb{S}$ is the group of finite permutations of $\mathbb{N}$.
- $\mathbf{A}_n = \mathbb{S}_n$ is the subgroup of permutations of $\{1, \ldots, n\}$.
- $f(X) = g(X_1)$ is a function $g : \mathbb{R} \to \mathbb{R}$ of the first entry.

Invariance is exchangeability:

$$X \;=\; (X_1, X_2, \ldots) \;\stackrel{\mathrm{d}}{=}\; (X_{\phi(1)}, X_{\phi(2)}, \ldots) \qquad \text{for each } \;\phi \in \mathbb{S}$$

$$X = (X_1,\; X_2,\; X_3,\; X_4,\; X_5, \ldots) \qquad\qquad \phi X = (X_3,\; X_1,\; X_2,\; X_4,\; X_5, \ldots)$$

$$f(X) = g(X_1) \qquad\qquad f(\phi X) = g(X_3)$$

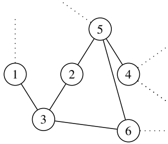
$$\tfrac{1}{|\mathbf{A}_n|} \textstyle\sum_{\phi \in \mathbf{A}_n} f(\phi X) \;=\; \tfrac{1}{n!} \textstyle\sum_{\phi \in \mathbb{S}_n} g(X_{\phi(1)}) \;=\; \tfrac{1}{n} \textstyle\sum_{i \leq n} g(X_i)$$

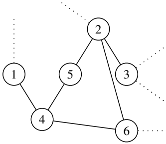
EXAMPLE: EXCHANGEABLE GRAPH

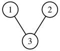
- X is the adjacency matrix of a random graph with vertex set $\mathbb{N}$.
- Again, $\mathbb{G} = \mathbb{S}$ and $\mathbf{A}_n = \mathbb{S}_n$.
- $f(X) = g((X_{ij})_{i,j=1,2,3})$ applies a function g to the first three (say) vertices.

If X is invariant, it is an exchangeable graph:

$$(X_{ij})_{i,j\in\mathbb{N}} \stackrel{\mathrm{d}}{=} (X_{\phi(i)\phi(j)})_{i,j\in\mathbb{N}} \qquad\qquad f(X) = g((X_{ij})_{i,j=1,2,3})$$





$f(X) = g\Bigg($  $\Bigg)$

$f(\phi X) = g($  $)$

$$\frac{1}{|\mathbb{S}_n|}\sum_{\phi\in\mathbb{S}_n} f(\phi X) \;=\; \frac{1}{n(n-1)(n-2)}\sum_{i,j,k\leq n \text{ distinct}} g(\text{subgraph of } X \text{ on } i,j,k)$$

Second Main Result

Central limit theorem (Austern & O.)

If $\mathbb{E}[|f(X)|^{2+\varepsilon}] < \infty$ and suitable conditions hold, then

$$\sqrt{|\mathbf{A}_n|}\Big(\int_{\mathbf{A}_n} f(\phi X) U_n(d\phi) - \mathbb{E}[f(X)|\text{ergodic component}]\Big) \quad \xrightarrow{\text{d}} \quad \eta Z$$

for $Z \sim N(0,1)$ and a positive random variable

$$\eta^2 = \int_{\phi \in \mathbb{G}} \mathbb{E}[f(X)f(\phi X)\,|\,\text{ergodic component}]\,|d\phi|$$

that is finite almost surely and independent of Z.

On the distribution: Invariance

$$X \stackrel{\text{d}}{=} \phi X \qquad \text{for all } \phi \in \mathbb{G}$$

On the group $\mathbb{G}$: Amenability

$$\frac{|\mathbf{A}_n \cap \psi \mathbf{A}_n|}{|\mathbf{A}_n|} \xrightarrow{n \to \infty} 1 \qquad \text{for finite/compact sets } \mathbf{A}_n \subset \mathbb{G} \text{ and all } \psi \in \mathbb{G}\ .$$

On the pair (f, X): Conditional mixing

If $M \subset \mathbb{G}$ is a set far away from a pair $\phi_, \phi_2 \in \mathbb{G}$, then

$$\big(f(\phi_1 X), f(\phi_2 X)\big) \perp\!\!\!\perp_{\text{ergodic component}} \big(f(\psi X)\big)_{\psi \in M} \qquad \text{approximately.}$$

Random fields

- Bolthausen's CLT: $X = (X_z)_{z\in\mathbb{Z}^d}$ stationary random field. Under mixing and moment condition,

$$\sqrt{n^d}\left(\frac{1}{n^d}\sum_{z\in\{0,\dots,n\}^d} f(X_z) - \mathbb{E}[f(X)|\text{erg. comp.}]\right) \quad \xrightarrow{d} \quad \eta Z \quad \text{as } n\to\infty$$

- Our results: $X = (X_t)_{t\in\mathbb{Z}^d}$ stationary random field. Under mixing and moment condition,

$$\sqrt{s^d}\ \left(\frac{1}{s^d}\int_{[0,s]^d} f(X_t)|dt| - \mathbb{E}[f(X)|\text{erg. comp.}]\right) \quad \xrightarrow{d} \quad \eta Z \quad \text{as } s\to\infty$$

Entropy of invariant processes

$$\sqrt{|\mathbf{A}_n|}\big(\text{empirical entropy}(n) - \text{entropy}\big) \quad \xrightarrow{d} \quad \eta Z \qquad \text{as } n\to\infty$$

Theorem (Austern & O.)
Let X be an exchangeable random structure and f a function with

$$\mathbb{E}[f(X)^2] < \infty \quad \text{and} \quad \sum_i \limsup_j \|f(X) - f(\tau_{ij}X)\|_{\mathbf{L}_2} < \infty\,.$$

Then

$$\frac{\sqrt{n}}{\eta}\Big(\tfrac{1}{n!}\sum_{\pi\in\mathbb{S}_n} f(\pi X) - \mathbb{E}[f(X)\,|\,\text{ergodic component}]\Big) \quad \xrightarrow{\text{d}} \quad Z \sim N(0,1)$$

The Wasserstein distance between the lhs and Z is

$$O\Big(\min_{k\in\mathbb{N}}\big[\tfrac{k^2}{\sqrt{n}} + \sum_{i>k}\limsup_j \|f(X) - f(\tau_{ij}X)\|_{\mathbf{L}_2}\big]\Big)\,.$$

For CLT aficionados
The asymptotic variance is $\eta^2 = \sum_{i,j\in\mathbb{N}} \mathrm{Cov}\big[\mathbb{F}^i(e), \mathbb{F}^j(\tau_{ij})\big|\mathbb{S}_\infty\big]$ where

$$\mathbb{F}^i(\phi) := \lim_{n\to\infty}\tfrac{1}{|\mathbb{S}_n^i|}\sum_{\phi'\in\mathbb{S}_n^i} f(\phi'\phi X) \qquad \text{and} \qquad \mathbb{S}_n^i := \{\phi\in\mathbb{S}_n \,|\, \phi(i) = i\}$$

random structure X	ergodic structures	CLT due to
exchangeable sequence	i.i.d. sequences	H. Bühlmann
exchangeable partition	paint-box dist.	
exchangeable graph	graphon distributions	Bickel, Chen, Levina (2011), Ambroise & Matias (2012)
jointly exch. array	dissociated array	Eagleson/Weber (1978)
separately exch. array	dissociated array	

Exchangeable graphs

$$\sqrt{n}\big(\text{estimator above} \;-\; \mathbb{E}[g(\text{subgraph of } X \text{ on } i,j,k)|\text{graphon}]\big) \;\xrightarrow{\text{d}}\; N(0,\eta)$$

Caron-Fox random graph model

$$\sqrt{n}\big(\text{empirical graphex subgraph density}(n) - \text{graphex subgraph density}\big) \;\xrightarrow{\text{d}}\; N(0,\eta)$$

Stochastic block models with a growing number of classes

Draw X_n from a stochstic block model with $m(n)$ classes, let $m(n)$ grow with n.

$$\frac{\sqrt{n}}{\eta_n}\big(\text{empirical triangle density}(n) - \text{population triangle density under } X_n\big) \;\xrightarrow{\text{d}}\; N(0,1)$$

Generalized U-statistics (Austern & O.)

- Elementary U-statistic: For $Y = (Y_1, Y_2, \ldots)$ and g function with k arguments,

$$\frac{(n-k)!}{n!}\sum_{\phi\in\mathbb{S}_n} g(Y_{\phi(1)},\ldots,Y_{\phi(k)}) \quad \text{or} \quad n^{-k}\sum_{i_1,\ldots,i_k\leq n} g(Y_{i_1},\ldots,Y_{i_k})$$

- Rewrite with $f\colon (y_i)\mapsto y_1$ coordinate function and $\mathbf{A}_n := \{0,1,\ldots,n-1\}$ as

$$n^{-k}\sum_{\phi_1,\ldots,\phi_k\in\mathbf{A}_n} g(Y_{1+\phi_1},\ldots,Y_{1+\phi_k}) = \frac{1}{|\mathbf{A}_n|^k}\sum_{\phi_1,\ldots,\phi_k\in\mathbf{A}_n} g(f(\phi_1 Y),\ldots,f(\phi_k Y))\ .$$

- On the right, substitute $\mathbb{Z}$ by an amenable group $\mathbb{G}$. We can still show asymptotic normality under suitable conditions.

THE RATE $\sqrt{\text{SOMETHING}}$

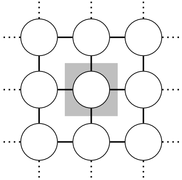
For i.i.d. variables
We usually encounter two rates:

- The "parametric" rate $\frac{1}{n}$
- The "nonparametric" rate $\frac{1}{\sqrt{n}}$

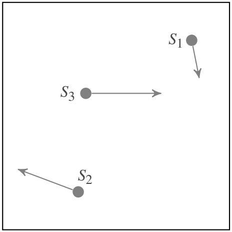
In our results

- In general, the rate is $\sqrt{|\mathbf{A}_n|}$.
- In the exchangeable case, we have $\sqrt{n}$.

Recall that we can think of $|\mathbf{A}_n|$ as sample size



AN ILLUSTRATION FROM PHYSICS



- Each particle s is described as $s = (x, y, z, \dot{x}, \dot{y}, \dot{z})$
- Model: n i.i.d. random particles $S_1, \ldots, S_n$
- Macroscopic observables (temperature, pressure, ...) are averages $A(S_1, \ldots, S_n)$
- If system is large enough, A is effectively constant

How large a system is "large enough"?

Thermodynamics:

- $n \approx 10^{23}$ ("thermodynamic limit").
- With $\sqrt{n}$ convergence, that is very large.

Random structures with dependence:

- More dependence $\to$ convergence may be slower than $\sqrt{\text{sample size}}$.
- Our results show: If symmetry (i.e. invariance) holds, we still get $\sqrt{\text{sample size}}$.

Example

The **Erdős-Rényi graph** with parameter $p \in [0, 1]$ has adjacency matrix

$$X = (X_{ij})_{i<j \in \mathbb{N}} \quad \text{with} \quad X_{ij} \sim_{\text{iid}} \text{Bernoulli}(p) \ .$$

Rates

- The convergence rate for the edge density in exchangeable graphs is $1/\sqrt{n}$.
- ER graphs are exchangeable, but the rate is known to be $1/n$.

Explanation

ER graphs have more symmetries: We are scaling by

- In an $n \times n$ exchangeable graph, we can permute n rows/columns.
- In an $n \times n$ ER graph, we can permute all $n(n+1)/2 \approx n^2$ entries above the diagonal.

More symmetries $\implies$ faster convergence.

UNIVERSALITY

Universality is the phenomenon that large-scale behavior is insensitive to fine details of the original problem.

Central limit theorem

If $X_1, X_2, \ldots$ i.i.d. (plus fine print), then

$$\sqrt{n}\Big(\frac{1}{n}\sum_{i\leq n} f(X_i) \;-\; \mathbb{E}[f(X_1)]\Big) \quad \xrightarrow{n\to\infty} \quad \text{normal}$$

Poisson limit theorem

If $X_{n1}, X_{n2}, \ldots$ i.i.d. for each n, and if $P(X_{n1} = 1) = \lambda/n$ for some $\lambda > 0$, then

$$\sum_{i\leq n} X_{ni} \quad \xrightarrow{n\to\infty} \quad \text{Poisson}(\lambda)$$

Wigner law

For random $n \times n$ matrix with i.i.d. entries,

$$\text{scaled distribution of eigenvalues} \xrightarrow{n\to\infty} \text{Wigner distribution}$$

Central limit theorem [Austern & O.]

For $\mathbb{G}$-invariant X and $f : \mathbf{X} \to \mathbb{R}$,

$$\sqrt{|\mathbf{A}_n|}\Big(\frac{1}{|\mathbf{A}_n|} \sum\nolimits_{\phi \in \mathbf{A}_n} f(X_i) \; - \; \mathbb{E}[f(X_1)] \Big) \quad \xrightarrow{n\to\infty} \quad Z$$

Poisson limit theorem [He, O. & Austern]

For $\mathbb{G}$-invariant X_n and $f : \mathbf{X} \to \{0, 1\}$, then

$$\sum\nolimits_{\phi \in \mathbf{A}_n} f(\phi X_n) \quad \xrightarrow{n\to\infty} \quad \text{compound Poisson}$$

Wigner law [Austern]

Morgane Austern has also obtained a Wigner law for matrices that are in a suitable sense $\mathbb{G}$-invariant.

It seems that (very loosely speaking) the intuition "i.i.d. implies universality" generalizes to "symmetry implies universality".

de Finetti and Ergodicity

Exchangeable sequences

$$(X_{\phi(1)}, X_{\phi(2)}, \dots) \quad \stackrel{\mathrm{d}}{=} \quad (X_1, X_2, \dots) \qquad \text{for } \phi \in \mathbb{S}$$

Here:

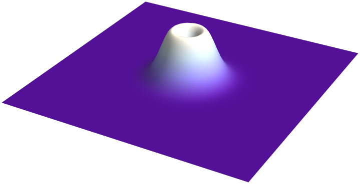
- We consider a random sequence $X = (X_1, X_2, \dots)$, say in $\mathbb{R}$.
- $\mathbb{S}$ is the group of all finite permutations of $\mathbb{N}$.
- P is the joint distribution of $(X_1, X_2, \dots)$.

de Finetti's theorem

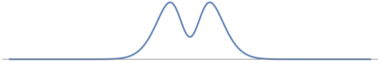
The set of exchangeable distributions on $\mathbb{R}^\infty$ is a convex set, and its extreme points are laws of i.i.d. sequences. A random sequence is exchangeable if and only if

$$P(X \in \bullet\,) \;=\; \int_{\mathcal{P}(\mathbb{R})} Q^{\otimes\infty}(X \in \bullet\,)\, \mu_P(dQ)$$

for some probability measure μ_P on the set $\mathcal{P}(\mathbb{R})$ of probability measures on $\mathbb{R}$.



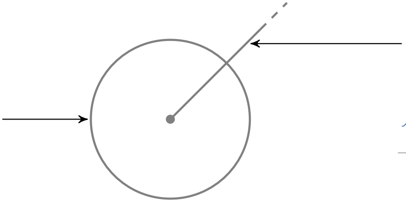
rotation-invariant distribution P on $\mathbb{R}^2$



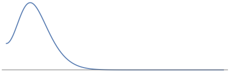
restriction to $\mathbb{R}$

$$P = \int_{\mathbb{R}_+} \eta_r \mu_P(dr) \qquad \text{where } \eta_r = \text{ uniform distribution on circle of radius } r$$

restriction of P to circle of radius r is η_r



μ_P = restriction of P to ray



Random $X \in \mathbb{R}^2$ rotation-invariant	**Random $X \in \mathbb{R}^\infty$ exchangeable**

Integral formulation

$$P(X \in \bullet) = \int_{\mathbb{R}_+} \eta_r(dx)\mu_P(dr)$$

where η_r is uniform on circle of radius r

$$P(X \in \bullet) = \int_{\mathcal{P}(\mathbb{R})} \eta_Q(dx)\mu_P(dQ)$$

where η_Q is the product distribution $Q^{\otimes\infty}$

Conditional formulation

X is conditionally uniform given its length.

X is conditionally i.i.d. given its marginal.

Generative formuluation

X can be generated by:

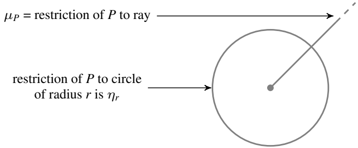
1. Generating a radius r at random.
2. Drawing X uniformly from the circle of radius r.

X can be generated by:

1. Select distribution Q at random.
2. Drawing $X_1, X_2, \ldots$ i.i.d. from Q.

Observation. A set is $\mathbb{G}$-invariant if and only if it is a disjoint union of orbits, i.e. a set of the form $\mathbb{G}(x) = \{\phi x \mid \phi \in \mathbb{G}\}$.

Simple case: Rotation invariance



- Each orbit is a circle.
- Each η_r is uniform on an orbit.

More complicated: Exchangeable sequences

- The orbits are countable. On infinite orbits, η_Q cannot be uniform.
- However, each η_Q concentrates on a distinct invariant set.

Property that holds for both

Each η concentrates on a union of orbits. In other words:

η is invariant and $\eta(A) \in \{0, 1\}$ for each $\mathbb{G}$-invariant Borel set A.

Setup

- $\mathbb{G}$ is a nice group.
  This essentially means $\mathbb{G}$ has a volume (formally: a Haar measure). All groups we have seen are nice.
- $\mathbb{G}$ acts measurably on $\mathbf{X}$.
  This means each $\phi \in \mathbb{G}$ defines a measurable map $\phi : \mathbf{X} \to \mathbf{X}$.

Ergodicity

A $\mathbb{G}$-invariant probability measure η is **ergodic** if

$$\eta(A) \in \{0, 1\} \qquad \text{for each Borel set } A \text{ with } \phi A = A \text{ for all } \phi \in \mathbb{G} .$$

Theorem (Farrell, Varadarajan).

The set of $\mathbb{G}$-invariant probability measures on $\mathbf{X}$ is a convex set. Its extreme points are the $\mathbb{G}$-ergodic measures, and any two ergodic measures are mutually singular. A probability measure P is $\mathbb{G}$-invariant if and only if

$$P = \int_{\mathcal{E}} \eta \, \mu_P(d\eta) .$$

Ergodic distributions

random structure X	invariance	ergodic distributions	characterizing theorem
random point in plain	rotations	uniform on circles	
sequence indexed by $\mathbb{N}$	exchangeability	i.i.d. distributions	de Finetti / Hewitt & Savage
graph with vertex set $\mathbb{N}$	exchangeability	graphon distributions	Aldous & Hoover
partition of $\mathbb{N}$	exchangeability	paintbox	Kingman
discrete-time Markov chain	stationarity	irreducible chains	
random field on $\mathbb{Z}^d$	stationarity		(sufficient condition: mixing)

Recall: Convergence

The ergodic decomposition $P = \int \eta \mu_P(d\eta)$ says we can sample P as:

1. Select ergodic component η from μ_P.
2. Draw $X \sim \eta$.

We observe a finite part of X as a sample.

$$\frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n} f(\phi X)|d\phi| \quad \xrightarrow{n\to\infty} \quad \mathbb{E}[f(X)|\,\text{ergodic distribution }\eta] \;=\; \mathbb{E}_\eta[f(X)]$$

$$P = \int_{\mathcal{E}} \eta \, \mu_P(d\eta)$$

This distribution generates the observed structure X.

This is the prior.

Statistical models

- A statistical model is a set $M \subset \mathcal{E}$ of ergodic measures.
  Each ergodic distribution reprsents one possible population.
- An (infinite) random object X is drawn from an ergodic measure η.
- A finite sample is "a small part" of X (say the first n entries if X is a sequence).

Bayesian interpretation

- The ergodic measure η is selected at random, according to μ_P.

Frequentist interpretation

- There is a true, fixed (but unknown) ergodic measure η that generates X.

Warning

The frequentist and Bayesian uses of the formalism imply very different null sets.

- Recall that ergodic measures are mutually singular: For each ergodic η, there is a measurable set A_η such that $\eta(A_\eta) = 1$
- Frequentist case: Suppose η_0 is the distribution of the data source. Then risks etc are computed under η_0. Here, $\eta_0(A_{\eta_0}) = 1$.
- Bayesian formalism: We select η randomly from μ_P. Risks are computed under $P = \int \eta d\mu_P$.
- Unless μ_P has an atom at η_0, we have $P(A_{\eta_0}) = 0$.

Explanation 1: Ergodicity

Rotations
Ergodic distributions are uniform distributions on orbits.

Permutations of infinite sequences

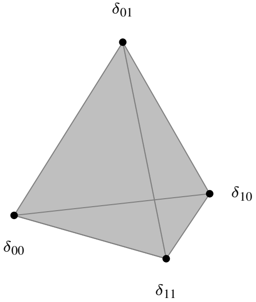
- Each orbit is either countably infinite or a singleton (if x is invariant).
- Any invariant distribution with countable support must be uniform on that support.

Since ergodic distributions are invariant, they must be point masses or have uncountable support.

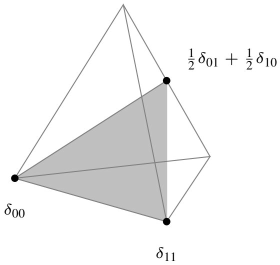
ONE WAY TO THINK ABOUT ERGODIC DECOMPOSITION

- We partition the space **X** into the orbits of $\mathbb{G}$.
- We "assign" each orbit to one ergodic measure η. (Up to tweaking null sets, there is only one way to do so.)
- The set A_η with $\eta(A_\eta) = 1$ is the union of all orbits assigned to η.
- Since the A_η are mutually disjoint, we cannot represent one ergodic measure as convex combinations (= mixtures!) of others.
- Since we cannot subdivide the A_η further (again up to null sets), every invariant measure is either ergodic or a mixture of ergodic measures.

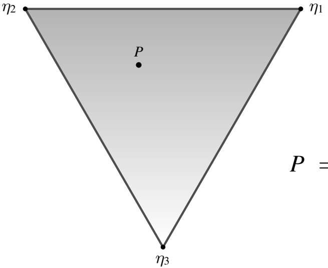
Explanation 2: The integral over extreme points.



The set of probability distributions on the set $\mathcal{X} = \{00, 01, 10, 11\}$.



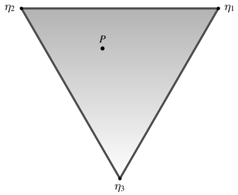
The set of exchangeable distributions on $\mathcal{X}$.



$P \; = \; \sum_{i=1}^{3} \mu_i \, \eta_i$ for convex coefficients μ_1, μ_2, μ_3

If the set of extreme points is not finite, the vector μ of convex coefficients becomes a probability measure.

Integral under a probability measure = generalized convex combination



A convex set K is a **simplex** if each point $y \in K$ can be represented as a (generalized) convex combination

$$y = \int_{\text{ex}\,K} z\,\mu_y(dz)$$

of the extreme points of K, and if μ_y is unique.

Some important examples of simplices:

- Compact convex metrizable set (by Choquet's theorem).
  The elements can be functions, operators, measures, etc.
- The set of distributions invariant under a nice group (ergodic decomposition).
- In particular, the set of exchangeabale distributions.
- The set of all distributions for a suitable sufficient statistic (Dynkin's theorem).

Some simplex results

de Finetti

$$P = \int Q^{\otimes\infty} \mu_P(dQ)$$

Bochner

A function $f : \mathbb{R} \to \mathbb{C}$ is positive definite iff

$$f(x) = \int_{\mathbb{R}} e^{-i\alpha x} \mu_f(d\alpha)$$

Bernstein

The functions $x \mapsto e^{-\alpha x}$ for $\alpha \geq 0$ are the extreme points of the set of completely monotonic functions on $\mathbb{R}_+$, and every such function is of the form

$$f(x) = \int_0^\infty e^{-\alpha x} c \cdot \mu_f(d\alpha) \qquad \text{where } c \geq 0$$

Lévy-Khinchine

A bounded completely monotonic function $f : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is infinitely divisible iff

$$-\log f(x) = \int_{\mathbb{R}} \frac{1 - e^{-\alpha x}}{1 - e^{\alpha}} c \cdot \mu_f(d\alpha) \qquad \text{where } c \geq 0$$

Summary I

result	conditions on the group	conditions on distribution
ergodic decomposition	lcscH	invariant
law of large numbers	lcscH + amenable	invariant + 1^{st} moment
central limit theorem	lcscH + amenable + metric	invariant + 2^{nd} moment + mixing
Berry-Esseen bound	lcscH + amenable + metric (+ geometry)	invariant + 4^{th} moment + mixing

A probability measure P is invariant iff it is conditionally ergodic iff

$$P = \int_{\mathcal{E}} \eta \, \mu_P(d\eta)$$

- The integral splits P into the information that is completely determined by a single instance of X (the ergodic measure η) and remaining information μ_P.
- Results like de Finetti, Aldous-Hoover, etc characterize the ergodic measures.
- A statistical model is a subset $M \subset \mathcal{E}$ of the ergodic measures.
- We can model η as fixed but unknown (frequentist) or as random (Bayesian).

result	conditions on the group	conditions on distribution
ergodic decomposition	lcscH	invariant
law of large numbers	lcscH + amenable	invariant + 1st moment
central limit theorem	lcscH + amenable + metric	invariant + 2nd moment + mixing
Berry-Esseen bound	lcscH + amenable + metric (+ geometry)	invariant + 4th moment + mixing

To recover statistics $\mathbb{E}[f(X)]$ of η, we average (approximately) over the group,

$$\frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n} f(\phi X)\,|d\phi|\;.$$

- This is a **symmetrization**. Recall the illustration from early on:

  We are learning/estimating by integrating out asymmetries.

- Sample averages, window estimates, and subgraph counts are special cases.
- Lindenstrauss' theorem shows (strong) consistency.
- Our results show universality (here: asymptotic normality).

Part II
Symmetrizing functions

I. Average over a finite group

If $\mathbb{G}$ is finite and f is a function on $\mathbb{R}^d$, a $\mathbb{G}$-invariant function can be constructed as

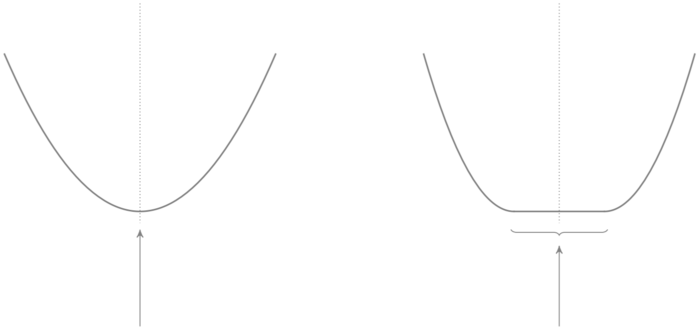
$$\bar{f}(x) \;=\; \frac{1}{|\mathbb{G}|} \sum\nolimits_{\phi\in\mathbb{G}} f(\phi x)$$

II. Construct a maximal invariant

III. Finite linear combination of invariant basis functions

IV. Represent function with a symmetry-preserving flow

V. Data augmentation



There is a single minimizer and this point is invariant.

The set of minimizers is invariant. It may not contain an invariant point.

If a function is invariant and has a unique minimizer, this minimizer is invariant.

Also true for minima that are local on a convex invariant neighborhood.

The summation trick

If $\mathbb{G}$ if finite or compact, we can construct a $\mathbb{G}$-invariant function by averaging.

$$\bar{f} \;=\; \frac{1}{|\mathbb{G}|}\sum_{\phi\in\mathbb{G}} f\circ\phi \qquad \text{or} \qquad \bar{f} \;=\; \frac{1}{|\mathbb{G}|}\int_{\mathbb{G}} f\circ\phi\,|d\phi|$$

Applications

- Used in machine learning to construct invariant/equivariant neural networks.
- Similarly, $\bar{P} = |\mathbb{G}|^{-1}\sum P\circ\phi^{-1}$ is an invariant probability measure.

Relation to amenability

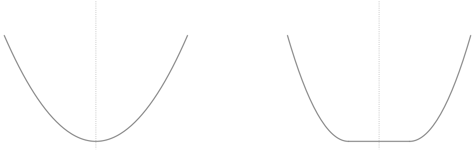
$$f_n \;:=\; \frac{1}{|\mathbf{A}_n|}\sum\nolimits_{\phi\in\mathbf{A}_n} f\circ\phi \qquad \text{or} \qquad f_n \;:=\; \frac{1}{|\mathbf{A}_n|}\int_{\mathbf{A}_n} f\circ\phi\,|d\phi|$$

If (!) the limit $\bar{f} = \lim_n f_n$ exists, amenability implies that is a $\mathbb{G}$-invariant function.

Relation to Part I

$$\int_{\mathbf{A}_n} f(\phi x)U_n(d\phi) \;=\; f_n(x)$$

$$Pf \;:=\; \frac{1}{|\mathbb{G}|}\sum_{\phi\in\mathbb{G}} f\circ\phi$$



If $\quad \ell$: function space $\to \mathbb{C} \quad$ is convex and $\mathbb{G}$-invariant, then

$$\ell(Pf) \;\leq\; \ell(f)$$

Note: Data augmentation makes ℓ approximately invariant.

$$Pf \;:=\; \frac{1}{|\mathbb{G}|}\sum_{\phi\in\mathbb{G}} f\circ\phi$$

Suppose $\mathbf{X}$ is a Hilbert space of functions whose inner product is $\mathbb{G}$-invariant,

$$\langle f\circ\phi, g\circ\phi\rangle \;=\; \langle f, g\rangle \;.$$

- The subset of $\mathbb{G}$-invariant functions is a norm-closed linear subspace.
- P is the orthogonal projector onto this subspace.
- If $A \subset \mathbf{X}$ is a convex $\mathbb{G}$-invariant subset, then $P(A) \subset A$.
- If H is an operator that commutes with all group elements, $H\circ\phi = \phi\circ H$, then

$$P(\text{eigenspace}) \;\subset\; \text{eigenspace}$$

Setup: Hypothesis test

- Two sets of distributions: H (the hypothesis) and A (the alternative)
- A test is given by a function $w : \mathbf{X} \to [0, 1]$ (the critical function), which we read as $w(x) =$ probability that test accepts if x is observed.

$$\hat{w} \in \arg\max_{w \in W} \inf \{P(w) \,|\, P \in A\} \qquad \text{subject to } \mathbb{E}_Q[w] \leq \alpha \quad \text{for all } Q \in H\,.$$

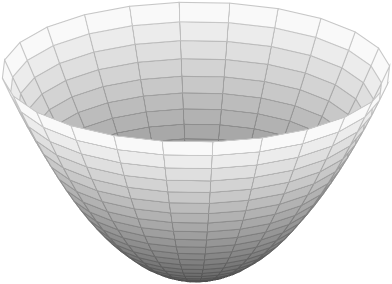
Hunt-Stein theorem

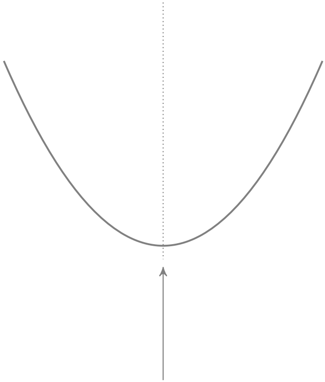
Consider two sets of distributions H (the hypothesis) and A (the alternative) that are both invariant as sets under an amenable group $\mathbb{G}$. If there is a critical function w that is minimax optimal at power α,

$$\hat{w} \in \arg\max_{w \in W} \inf \{P(w) \,|\, P \in A\} \qquad \text{subject to } \mathbb{E}_Q[w] \leq \alpha \quad \text{for all } Q \in H\,.$$

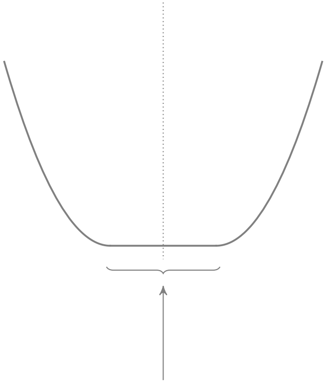
there is also a $\mathbb{G}$-invariant critical function that is minimax-optimal at power α.

Suppose $f : V \to \mathbb{R} \cup \{\infty\}$ is a convex function. If f has a minimizer v, does it also have an invariant minimizer $\bar{v}$?





$\arg\min f$ is a single point
this point is invariant



$\arg\min f$ is an invariant set
in this example, it contains an invariant point,
but that need not be true in general

Invariant points

- Groups that include shifts (like the crystallographic groups) have no invariant points.
- The rotation group of $\mathbb{R}^d$ has one invariant point (the center of rotation).

Invariant functions

- Invariant functions always exist (the constant functions).
- Non-trivial shift-invariant function *may* exist: True e.g. for discrete shifts on $\mathbb{R}$.

VAGUE INTUITION

Consider a group of transformations of some space Ω.

$$\begin{tikzcd}
 & \text{points in } \Omega \arrow[dl] \arrow[dr] & \\
\text{functions on } \Omega \arrow[dr] & & \text{distributions on } \Omega \arrow[dl] \\
 & \text{distributions of functions on } \Omega &
\end{tikzcd}$$

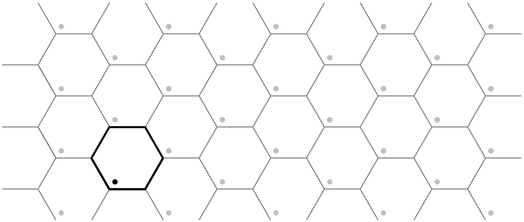
(i.e. random fields, stochastic processes etc)

$\downarrow$ more invariant objects

larger space $\Longrightarrow$ symmetry becomes a weaker constraint

We will now look at one specific class of (amenable) groups.

A group of isometries (rigid motions) of $\mathbb{R}^d$ is **crystallographic** if it tiles $\mathbb{R}^d$ with a convex polytope.



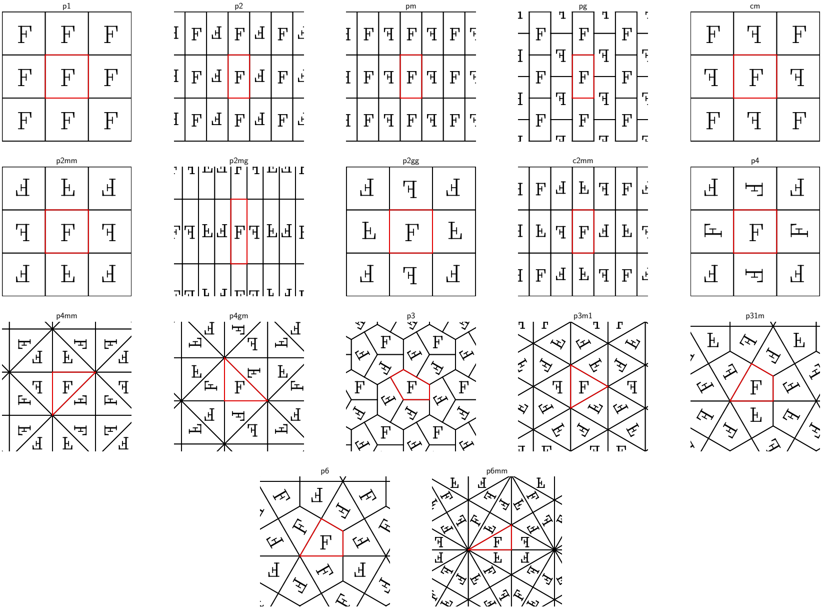
Classification

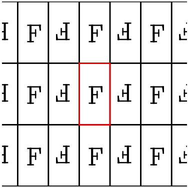
- Crystallographic groups are amenable: We can fix any point $x_0 \in \mathbb{R}^d$ and set

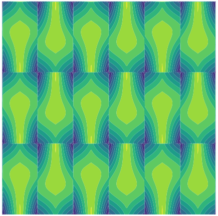
$$\mathbf{A}_n = \{\phi \in \mathbb{G} \mid \phi x_0 \text{ is in ball of radius } n \text{ around } x_0\} .$$

- There are 17 crystallographic groups on $\mathbb{R}^2$, and 230 on $\mathbb{R}^3$.

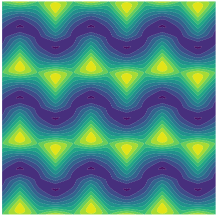
THE 17 CRYSTALLOGRAPHIC GROUPS ON THE PLANE







invariant but not continuous



invariant and continuous

Method 0: Data augmentation

Approximate, randomized symmetrization of the empirical measure of a sample.

Method I: Use the summation trick

Method II: Construct a maximal invariant

A maximal invariant is a function $\rho : \mathbb{R}^d \to M$ such that

$$\rho(x) \;=\; \rho(y) \qquad \Longleftrightarrow \qquad x \;=\; \phi y \text{ for some } \phi \in \mathbb{G}$$

If $h : M \to \mathbb{R}$ is a function, then $f = h \circ \rho$ is a $\mathbb{G}$-invariant function $\mathbb{R}^d \to \mathbb{R}$.

Caution: Only some group actions have (measurable) maximal invariants.

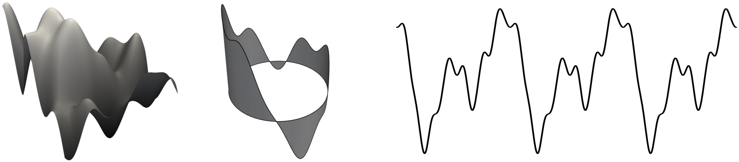
Method III: Find a basis of invariant functions

If $\xi_1, \xi_2, \ldots$ are a basis of $\mathbb{G}$-invariant functions, each function of the form

$$f \;=\; \sum_{n \in \mathbb{N}} c_n \xi_n \qquad \text{for } c_1, c_2, \ldots \in \mathbb{R}$$

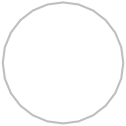
is $\mathbb{G}$-invariant (under suitable regularity conditions).

MAXIMAL INVARIANTS FOR CRYSTALLOGRAPHIC GROUPS



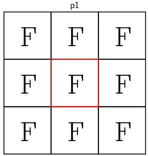
function h on $\mathbb{R}^2$ restrict h to circle unfold circle and replicate

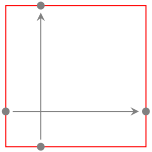
Each point x on the line is represented by exactly one point $\rho(x)$ on the circle.

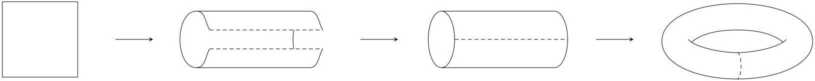
$\rho : \mathbb{R} \longrightarrow$ 

$h \circ \rho =$ periodic function on $\mathbb{R}$

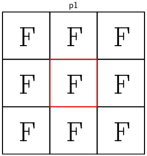
GLUING



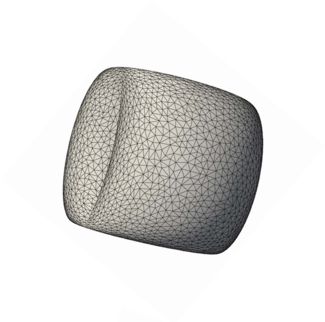


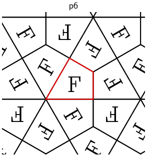


1) Pair faces of polytopes according transformations in group.
2) Glue paired faces to each other.
3) Embed the resulting surface in $\mathbb{R}^D$ with multidimensional scaling.

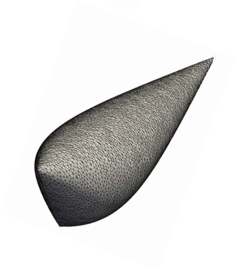


$\rho : \mathbb{R}^2 \longrightarrow$

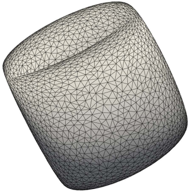




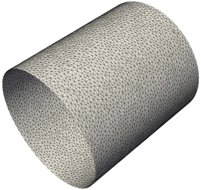
$\rho : \mathbb{R}^2 \longrightarrow$



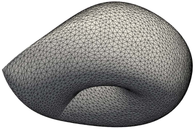
p1

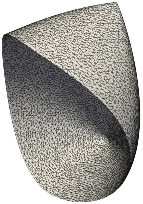


pm

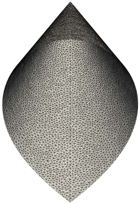


p2gg

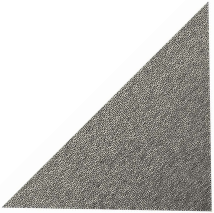




p2mg



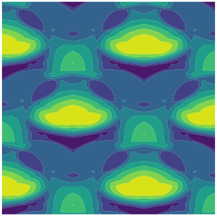
p4gm

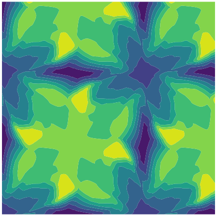


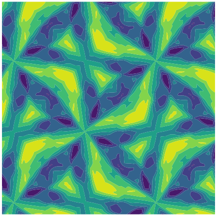
p3m1

INVARIANT NEURAL NETWORKS

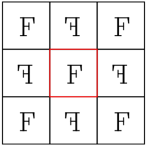
neural network on $\mathbb{R}^D \circ \rho \quad = \quad$ invariant neural network on $\mathbb{R}^2$



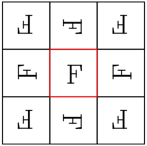




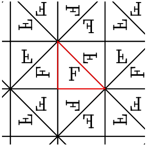
cm



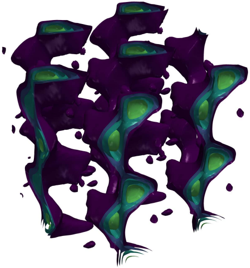
p4

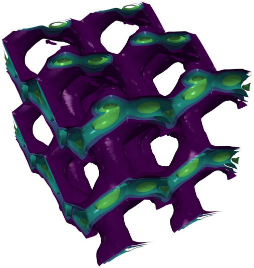


p4gm



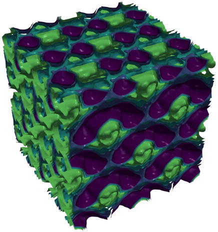
$P3_2$

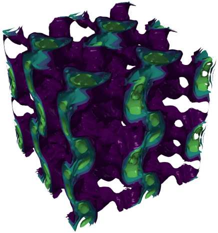




Pnn2

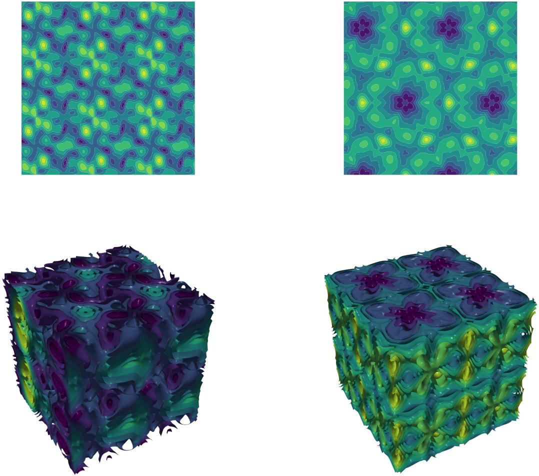
Aea2





$P4_12_12$

$$\text{GP path } \mathbb{R}^D \circ \rho \quad = \quad \text{invariant GP path on } \mathbb{R}^2$$



INVARIANT BASES

Fourier Theorem

If $\mathbb{G}$ is a crystallographic group on $\mathbb{R}^d$, the constrained partial differential equation

$$\begin{aligned} -\Delta e &= \lambda e \\ e &= e \circ \phi \qquad \text{for all } \phi \in \mathbb{G} \end{aligned}$$

has solutions for a countably many values of λ. These values satisfy

$$0 = \lambda_1 \leq \lambda_2 \leq \ldots \quad \text{and} \quad \lambda_n \xrightarrow{n\to\infty} \infty .$$

There is a sequence of eigenfunctions $e_1, e_2, \ldots$ that forms an orthonormal basis of $\mathbf{L}_2$(polytope). A continuous function f on $\mathbb{R}^d$ is $\mathbb{G}$-invariant if and only if

$$f = \sum_{n=1}^{\infty} c_n e_n \qquad \text{for some } c_1, c_2, \ldots \in \mathbb{R} ,$$

where the series converges in the supremum norm.

We can represent every invariant function in a basis of invariant functions e_i that generalize the Fourier basis.

In one dimension, the Laplacian Δ is the second derivative $\Delta = d^2/dx^2$. Recall:

$$(-\Delta)\sin(\nu x) \;=\; \nu^2\sin(\nu x) \quad \text{and} \quad (-\Delta)\cos(\nu x) \;=\; \nu^2\cos(\nu x)$$

That is just the eigen-equation $(-\Delta f) = \lambda f$.

> $\sin(\nu x)$ and $\cos(\nu x)$ are eigenfunctions of $-\Delta$. The eigenvalues are the quadratic frequencies $\lambda = \nu^2$.

In other words, the Fourier transform is a change of basis to an ONB of sines/cosines.

Δ is self-adjoint on a space $\mathcal{F} \subset \mathbf{L}_2(\Pi)$ of functions if we can swap f and h in

$$\langle -\Delta f, h \rangle_{\mathbf{L}_2} = \int_\Pi f(x)h(x)\mathrm{vol}_n(dx) \qquad \text{for } f, h \in \mathcal{F} \ .$$

Green identity. Under suitable conditions on Π, the Laplacian satisfies

$$\langle -\Delta f, h \rangle_{\mathbf{L}_2} = \int_{\Pi^\circ} \nabla f(x)^t \nabla h(x)\mathrm{vol}_n(dx) - \int_{\partial\Pi} \nabla f(x)^t \mathbf{N}_\Pi(x) h(x)\mathrm{vol}_{n-1}(dx) \ .$$

Flux theorem (Adams & O., 2023).

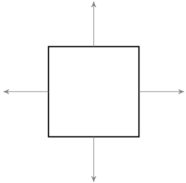
If $\mathbb{G}$ is crystallographic and $F : \Pi \to \mathbb{R}^n$ satisfies

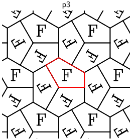
$$F(\phi x) = (\text{orthogonal component of } \phi) \cdot F(x)$$

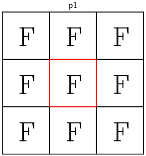
its integral normal to the boundary cancels,

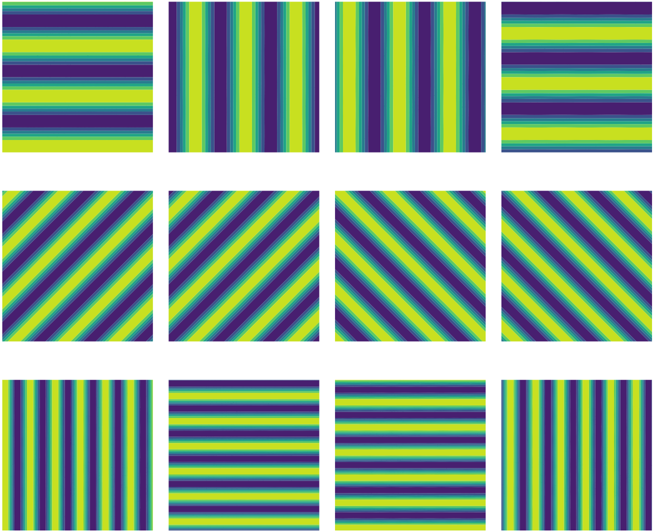
$$\int_{\partial\Pi} F(x)^t \mathbf{N}_\Pi(x)\mathrm{vol}_{n-1}(dx) = 0 \ .$$

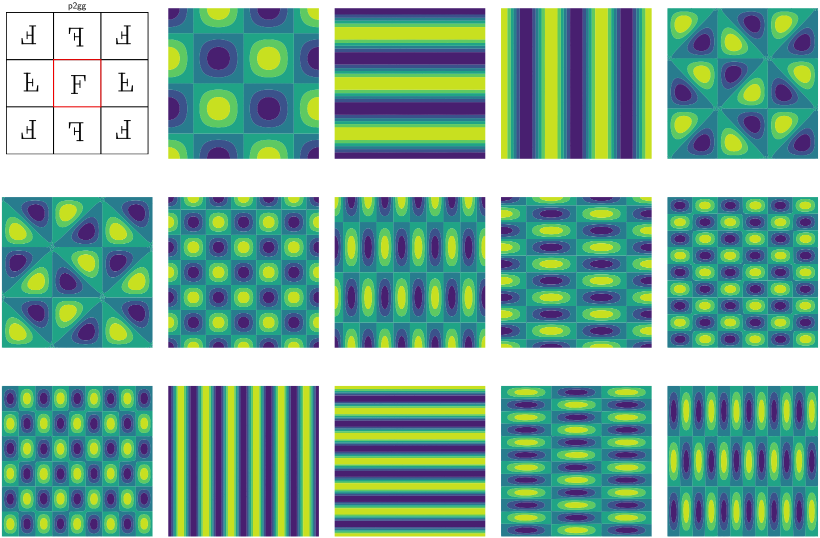
That holds in particular for $F = \langle \nabla f, \mathbf{N} \rangle h$.

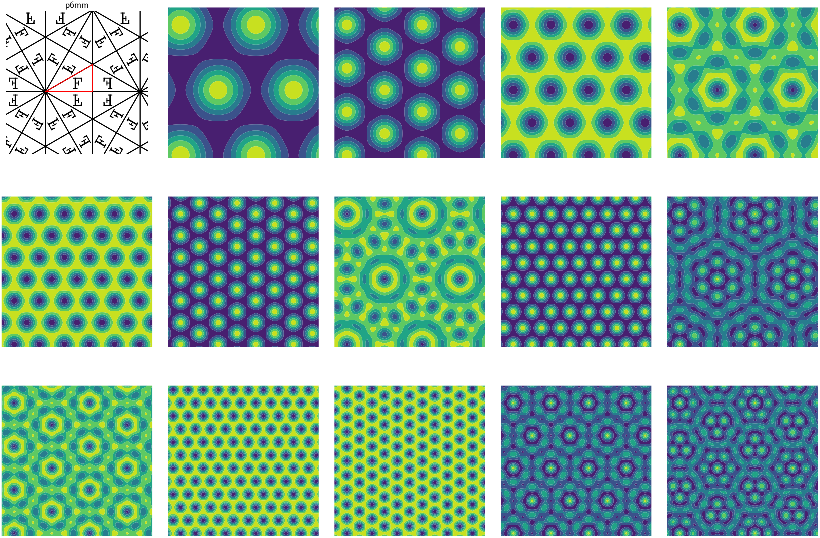




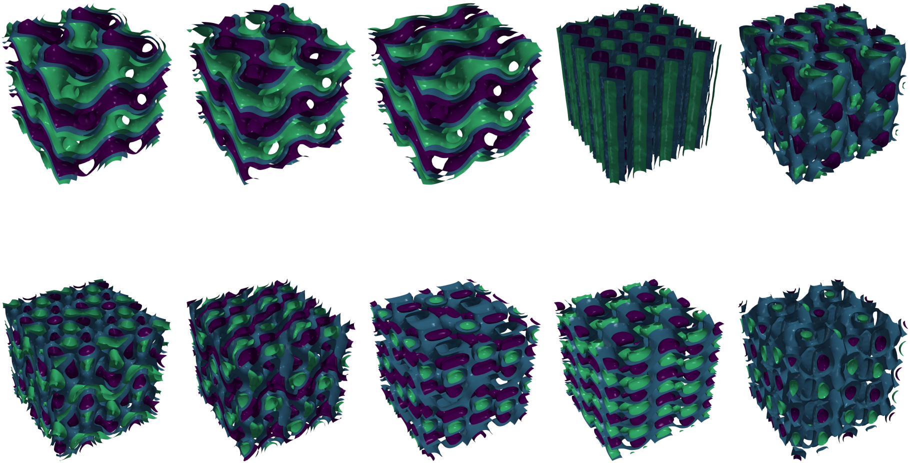

p1
F
F
F
F
F
F
F
F
F

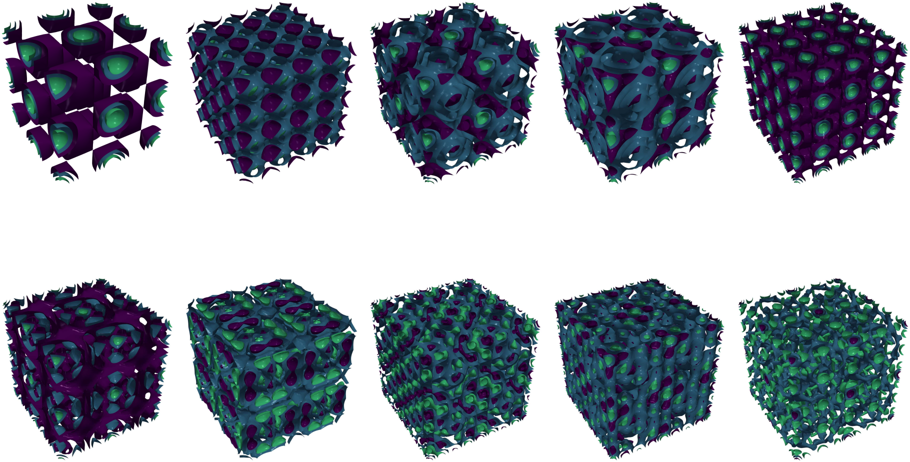



p2gg


p6mm

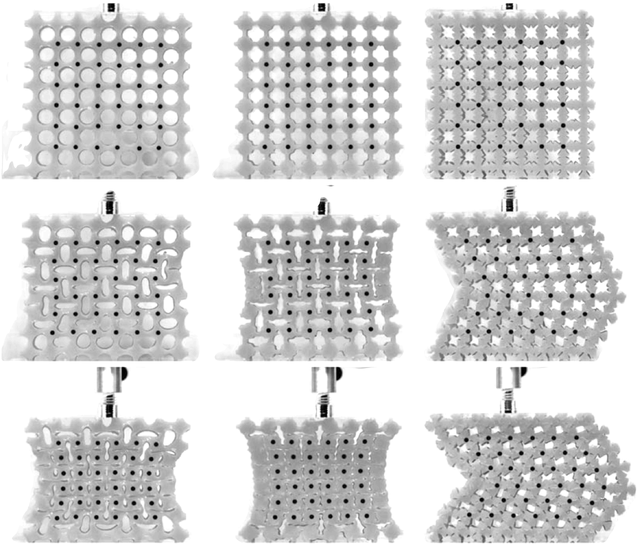
FOURIER BASES ON $\mathbb{R}^3$

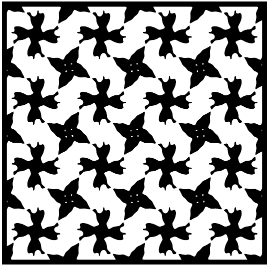




TOPOLOGICAL OPTIMIZATION

I will make the slides on meta-materials available once the preprint is public.





$$\text{shape} \quad = \quad \text{function} \quad s : [0,1]^2 \rightarrow \{0,1\}$$

$$\text{fixed object} \xrightarrow[\text{parameterized by } \theta]{\text{property-preserving transformation}} \text{object parameterized by } \theta$$

Examples

$$\text{fixed probability measure} \xrightarrow[\text{parameterized by } \theta]{\text{normalizing flow}} \text{distribution parameterized by } \theta$$

$$\text{Euclidean volume} \xrightarrow[\text{parameterized by } \theta]{\text{pointwise rewaiting by density}} \text{distribution parameterized by } \theta$$

Here:

$$\text{fixed symmetric shape} \xrightarrow[\text{parameterized by } \theta]{\text{symmetry-preserving flow}} \text{shape parameterized by } \theta$$

SYMMETRY-PRESERVING FLOWS

Every element of a crystallographic group is an isometry, and hence of the form

$$\phi(x) \;=\; A_\phi x + b_\phi \qquad \text{for an orthogonal matrix } A_\phi \text{ and } b_\phi \in \mathbb{R}^d \;.$$

Theorem sketch
Consider a function $H : \mathbb{R}^d \to \mathbb{R}^d$ that satisfies

$$A_\phi H = H \circ \phi \qquad \text{for each } \phi \in \mathbb{G} \;. \tag{$*$}$$

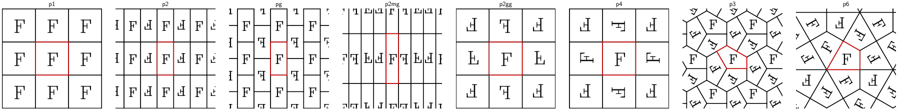
If a smooth function H satisfies $(*)$, the solution F of

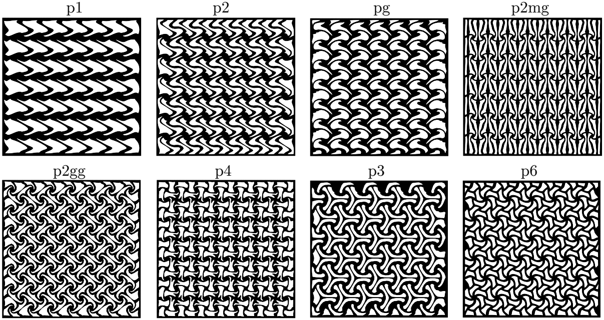
$$\frac{d}{dt}F(x,t) \;=\; H(F(x,t)) \quad \text{and} \quad F(x,0) = x$$

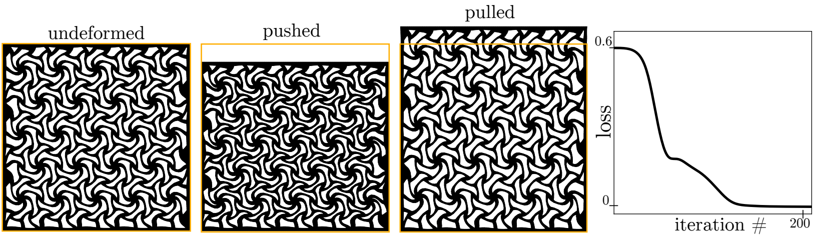
is a flow that preserves $\mathbb{G}$-invariance.

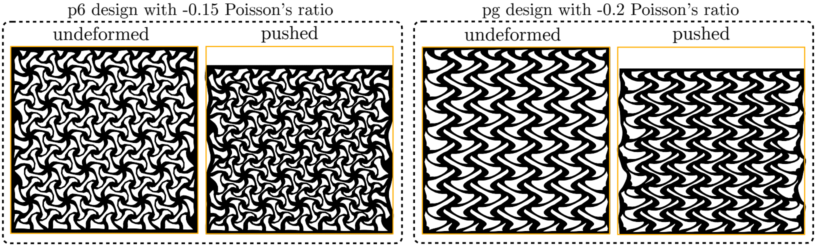
Method sketch

- Fix a $\mathbb{G}$-invariant shape s_0.
- Use our symmetrization tools to construct a neural network H_θ that satisfies $(*)$.
- That gives us a symmetry-preserving flow F_θ.
- Define the class of symmetric shapes $\{s_\theta = \text{deformation of } s_0 \text{ by } F_\theta \,|\, \theta \in \mathcal{T}\}$.
- Combine this with a standard optimizer for metamaterial properties.

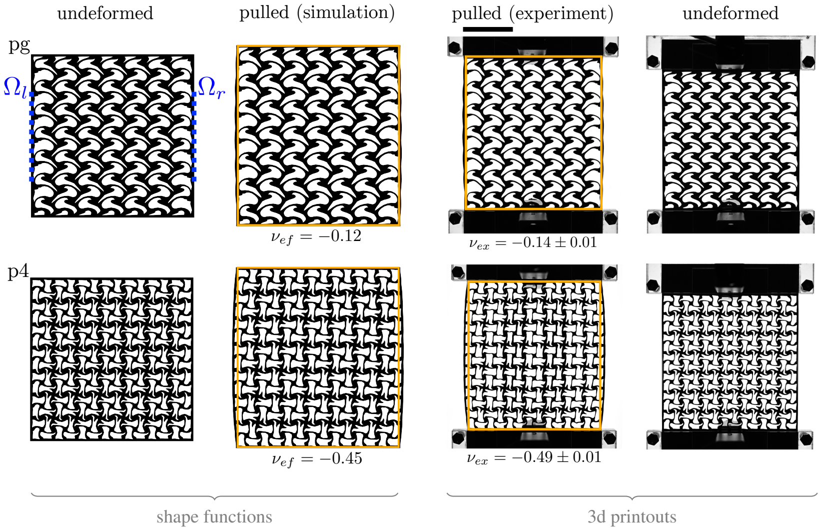

p1
p2
pg
p2mg
p2gg
p4
p3
p6


p1
p2
pg
p2mg
p2gg
p4
p3
p6


undeformed
pushed
pulled
0.6
loss
0
iteration #
200


p6 design with -0.15 Poisson's ratio
undeformed
pushed
pg design with -0.2 Poisson's ratio
undeformed
pushed

RESULTS


undeformed
pulled (simulation)
pulled (experiment)
undeformed
pg
Ω_l
Ω_r
$\nu_{ef} = -0.12$
$\nu_{ex} = -0.14 \pm 0.01$
p4
$\nu_{ef} = -0.45$
$\nu_{ex} = -0.49 \pm 0.01$
shape functions
3d printouts

REFERENCES

The reference on the pointwise ergodic theorem:

E. Lindenstrauss, *Pointwise theorems for amenable groups*, Invent. Math., 146(2)259–295, 2001.

On the connection between the pointwise theorem and statistics, and the various central limit theorems, see:

M. Austern and P. Orbanz, *Limit theorems for distributions invariant under groups of transformations*, Ann. Statist. 50(4)1960–1991, 2022.

Other types of universality (Poisson approximation and Wigner):

H. Ye, P. Orbanz and M. Austern, *Poisson approximation for stochastic processes summed over amenable groups*, arxiv 2401.10060.
M. Austern, *A free central limit theorem for dynamical systems*, arxiv 2005.10923

The general statement of the ergodic decomposition theorem is due to:

V. R. Varadarajan, *Groups of authomorphisms of Borel spaces*, Trans. Amer. Math. Soc., 109(2)191–220, 1963.

A good reference on ergodic decomposition is also appendix A.1 of Kallenberg's book (see Theorem A1.4, although this version assumes the group is countable):

O. Kallenberg, *Probabilistic symmetries and invariance principles*, Springer 2005.

On the relationship between ergodic decomposition and de Finetti, Aldous-Hoover, Kingman etc, see:

P. Orbanz and D. M. Roy, *Bayesian models of graphs, arrays and other exchangeable random structures*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 37, 437-461, 2015.

The maximal invariants for crystallographic groups are from:

R. P. Adams and P. Orbanz, *Representing and Learning Functions Invariant Under Crystallographic Groups*, arxiv 2306.05261.

Topological optimization and meta-materials:

M. Mirramezani, A. S. Meeussen, K. Bertoldi, P Orbanz, and R. P. Adams, *Designing Mechanical Meta-Materials by Learning Equivariant Flows.*, ICLR 2025.