Repr. theory, Fourier analysis, invariants

Risi Kondor, The University of Chicago

Black box ML

Data: $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$

$\downarrow$

} algorithm

$\downarrow$

Hypothesis: $\widehat{f}\colon \mathbf{x} \mapsto y$

Features

$$\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_m, y_m)\}$$

$$\downarrow$$

$$\{(\boldsymbol{\phi}(\mathbf{x}_1), y_1), \ldots, (\boldsymbol{\phi}(\mathbf{x}_m), y_m)\}$$

$$\downarrow$$

} algorithm

$$\downarrow$$

Hypothesis: $\widehat{f} \colon \boldsymbol{\phi}(\mathbf{x}) \mapsto y$

The individual coordinates $(\phi_1(\mathbf{x}), \ldots, \phi_n(\mathbf{x}))$ of $\boldsymbol{\phi}(\mathbf{x})$ are called **features**.

Regularization

Many algorithms are, in fact, linear in the feature space, i.e.,

$$f(\boldsymbol{\phi}(\mathbf{x})) = \sum_i w_i \, \phi_i(\mathbf{x})$$

and what we really learn is $\mathbf{w} = (w_1, \ldots, w_n)$. Regularization:

- ℓ_2 –regularization: $\Omega(f) = \|\mathbf{w}\|_2^2 = \sum_i |w_i|^2$
- ℓ_1 –regularization: $\Omega(f) = \|\mathbf{w}\|_1 = \sum_i |w_i|$

In both cases, the choice of features is critical. In physics applications, it is also important that the features be **invariant**.

Classical invariants

Let G be a group acting on a vector space $\mathcal{X}$. A function $\Upsilon\colon \mathcal{X} \to \mathbb{R}$ is said to be an **invariant** to the action of G if

$$\Upsilon(g(\mathbf{x})) = \Upsilon(\mathbf{x}) \quad \text{for all} \quad x \in \mathcal{X}, \;\; g \in G.$$

Example:

- $\mathrm{SO}(3)$ acting on $\mathbb{R}^3$ and $\Upsilon(\mathbf{x}) = \|\mathbf{x}\|$.

Actually, we are more interested in the invariants of functions.

Let G be a group acting on a vector space $\mathcal{X}$ and let V be a space of functions on $\mathcal{X}$ (e.g., $V = L_2(\mathcal{X})$).

The action of G on $\mathcal{X}$ extends naturally to V by

$$f \mapsto f' = g(f) \qquad f'(x) = f(g^{-1}(x)).$$

A function $\Upsilon : V \to \mathbb{R}$ is said to be an **invariant** w.r.t. the action of G if

$$\Upsilon(g(f)) = \Upsilon(f) \quad \text{for all} \quad f \in V, \;\; g \in G.$$

Examples for the case of translations acting on $\mathbb{R}$:

- $\Upsilon_0 = \int |f(x)|^2 dx$
- $\Upsilon_\omega = |\widehat{f}(\omega)|^2$ for any frequency ω.

1: Autocorrelation (for translations)

The **autocorrelation** of f is

$$a(x) = \int f(x+y)f(y)dy.$$

Tells us how much f changes when we translate it by an amount y. Clearly invariant to translation (assuming $f'(x) = f(x-g)$):

$$\int f'(x+y)f'(y)dy = \int f(x+y)f(y)dy.$$

2: Power spectrum (for translations)

The **power spectrum** of f is

$$\widehat{a}(\omega) = \widehat{f}(\omega)^* \cdot \widehat{f}(\omega) = |\widehat{f}(\omega)|^2.$$

Literally measures the amount of energy in each Fourier mode. Clearly invariant to translation:

$$\widehat{a}_{f'}(\omega) = (e^{2\pi i \omega g} \widehat{f}(\omega))^* \cdot (e^{2\pi i \omega g} \widehat{f}(\omega)) = \widehat{f}(\omega)^* \cdot \widehat{f}(\omega) = \widehat{a}_f(\omega).$$

- The power spectrum is just the Fourier transform of the autocorrelation. Limitation of both: they lose all the “phase information” in f.

3: Triple correlation & bispectrum

The **triple correlation** of f is

$$b(x_1, x_2) = \int f(y - x_1)\, f(y - x_2)\, f(y)\, dy$$

The **bispectrum** of f is

$$\widehat{b}(k_1, k_2) = \widehat{f}(k_1)^* \, \widehat{f}(k_2)^* \, \widehat{f}(k_1 + k_2).$$

- Again, these are both invariants, and the bispectrum is the Fourier transform of the triple correlation. Obviously, they are highly redundant (overcomplete).

Reconstructing f from b

$$\widehat{b}(k_1, k_2) = \widehat{f}(k_1)^* \, \widehat{f}(k_2)^* \, \widehat{f}(k_1 + k_2).$$

Use the following algorithm to recover f from $\widehat{b}$:

1. $\widehat{f}(0) = \widehat{b}(0,0)^{1/3}$
2. $\widehat{f}(1) = e^{i\phi}\sqrt{\widehat{b}(0,1)/\widehat{f}(0)}$ $\rightarrow$ indeterminacy in ϕ inevitable
3. $$\widehat{f}(k+1) = \frac{\widehat{b}(1,k)}{\widehat{f}(1)^* \, \widehat{f}(k)^*} k = 2, 3, \ldots$$

Conclusion: If $\widehat{f}(k) \neq 0$ for any k, then $\widehat{b}$ uniquely determines $\widehat{f}$ up to translation, i.e., the bispectrum is **complete**.

Is there a general theory behind all this? In particular, is there a natural generalization of Fourier analysis to groups other than $\mathbb{R}^d$?

Groups

Groups

A set G with a binary operations $\cdot : G \times G \to G$ is called a **group** if

1. for any $g_1, g_2 \in G$, $g_2 g_1 \in G$ (closure)
2. for any $g_1, g_2, g_3 \in G$, $g_3(g_2 g_1) = (g_3 g_2) g_1$ (associativity)
3. $\exists e \in G$ such that $e g = g e = g$ for any $g \in G$ (identity)
4. For any $g \in G$ there is a $g^{-1} \in G$ such that $g^{-1} g = e$ (inverse).

Groups play a fundamental role in Physics because they are the natural algebraic structure to describe invariances. $\big[$ G is said to be **commutative** or **Abelian** if $g_1 g_2 = g_2 g_1$ for all $g_1, g_2 \in G$ $\big]$.

Examples of countable groups

- The cyclic groups $\mathbb{Z}_n = 0, 1, \ldots, n-1$ (addition modulo n).
- Klein's Viergruppe $V = \{1, i, j, k\}$.
- Quaternion group $Q = \{1, i, j, k, -1, -i, -j, -k\}$.
- Icosahedron group $I_h \equiv A_5$.
- Symmetric groups $\mathbb{S}_n$ (group of permutations).
- The integers $\mathbb{Z}$.

Examples of continuous groups

- The reals $\mathbb{R}$ and the Euclidean spaces $\mathbb{R}^d$.
- The rotation groups $\mathrm{SO}(n)$.
- The Euclidean group $\mathrm{ISO}(n)$ and the rigid body motions $\mathrm{ISO}^+(n)$.
- The special unitary groups $\mathrm{SU}(n)$.
- The Lorentz group $\mathrm{SO}(3,1)$.
- The general linear group $\mathrm{GL}(n)$.

Types of continuous groups

- non-Lie groups
- Lie groups
  - Compact Lie groups (e.g., $\mathrm{SO}(n)$, $\mathrm{SU}(n)$)
  - Non-compact Lie groups (e.g., $\mathbb{R}$, $\mathrm{SO}(3,1)$, $\mathrm{GL}(n)$ etc.)

Many of these groups can be thought of as subsets of $\mathrm{GL}(n)$.

Compact groups are nice for many reasons, including the fact that they have a uniquely defined invariant measure μ, called the Haar measure.

Group actions

The **action** of a group G on a set $\mathcal{X}$ is a collection of mappings

$$g\colon \mathcal{X} \to \mathcal{X} \qquad g \in G$$

such that

$$(g_2 g_1)(x) = g_2(g_1(x)) \qquad \forall g_1, g_2 \in G.$$

The action is said to be **transitive** if for any $x, x' \in \mathcal{X}$

$$\exists g \in G \quad \text{such that} \quad g(x) = x'.$$

Erlagen program: Geometry is the study of properties invariant under a group (Felix Klein, 1872).

We are particularly interested in the actions of groups on vector spaces.

- If G acts on $\mathcal{X}$ and V is an (invariant) vector space of functions on $\mathcal{X}$, then we have the natural induced action on V

$$T_g \colon f \to f' \qquad f'(x) = f(g^{-1}(x)).$$

Key question: How does V fall apart into a direct sum of subspaces that are invariant (fixed) under all the T_g's?

Representations

Representations

$$\begin{array}{ccc} \{g_1,\, g_2\} & \longrightarrow & g_2 \cdot g_1 \\ \big\downarrow & & \big\downarrow \\ \{\rho(g_1),\, \rho(g_2)\} & \longrightarrow & \rho(g_2) \cdot \rho(g_1) \end{array}$$

Given a group G and a vector space V (over $\mathbb{C}$), a collection of invertible operators $\{\rho(g)\}_{g \in G}$ on V is a **representation** of G if

$$\rho(g_2) \cdot \rho(g_1) = \rho(g_2 g_1)$$

for all $g_1, g_2 \in G$.

This is just an action of G on V realized via the $\rho(g)$ operators. (But is it transitive?) Equivalently, $\rho \colon G \to \mathrm{GL}(V)$ is a homomorphism.

Example: representations of Q

Recall that the quarterion group $Q = \{1, i, j, k, -1, -i, -j, -k\}$ is defined by

$$i^2 = j^2 = k^2 = -1, \qquad (-1)a = -a \qquad ij = k.$$

One represention of Q:

$$\rho(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \rho(i) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$$

$$\rho(j) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \qquad \rho(k) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

so $\{\rho_\omega\}_{\omega\in\mathbb{R}}$ are representations of $\mathbb{R}$.

Unitary representations

Typically, V is a Hilbert space, and we can talk about unitary representations.

A representation $\rho\colon G \to V$ is **unitary** if each of the $\rho(g)$ operators are unitary, i.e.,

$$\langle \rho(g)(x), \rho(g)(y) \rangle = \langle x, y \rangle$$

for all $g \in G$ and all $x, y \in \mathcal{X}$. Equivalently, $(\rho(g))^{-1} = (\rho(g))^{\dagger}$.

In the following we will deal almost exclusively with unitary representations.

How many representations does a given group have? What are they?

Equivalence

$$\begin{array}{ccc} V_1 & \xrightarrow{\rho_1} & V_1 \\ T\big\downarrow & & T\big\downarrow \\ V_2 & \xrightarrow{\rho_2} & V_2 \end{array}$$

Let $\rho_1 \colon G \to V_1$ and $\rho_2 \colon G \to V_2$ be two representations of G. The two representations are said to be **equivalent**, denoted $\rho_1 \cong \rho_2$, if there is some fixed bijection $T \colon V_1 \to V_2$ such that

$$T^{-1} \circ \rho_2(g) \circ T = \rho_1(g) \qquad \forall g \in G.$$

Equivalent representations are often considered the same.

Reducible representations

If W has a non-trivial subspace of V such that

$$\rho(x) \in W \qquad \forall x \in W,$$

then ρ is said to be **reducible**. Otherwise it is **irreducible**. (The irreducible representations of commutative groups are always one dimensional)

Obviously, in this case $\rho\downarrow_W$ is also a representation. But is $\rho\downarrow_{W^\perp}$ also a representation?

Complete reducibility

Theorem. Let ρ be a representation of a *compact* group G on a Hilbert space V over $\mathbb{C}$. Then if W is an invariant subspace of V, then its orthogonal complement, $W^\top$ is also an invariant subspace.

Corollary. ρ decomposes into the direct sum of representations

$$\rho = \rho_W \oplus \rho_{W^\perp}.$$

Complete reducibility

Corollary. Let $\mathcal{R}$ be a **complete set of inequivalent irreducible representations ("irreps")** of a compact group G. Then any representation μ of G can be uniquely expressed in the form

$$\mu = \bigoplus_{\rho \in \mathcal{R}} \bigoplus_{i=1}^{\kappa_\mu(\rho)} \rho = \bigoplus_{\rho \in \mathcal{R}} \rho^{\oplus \kappa_\mu(\rho)}.$$

The irreps are the "primes" in the world of representations of compact groups.

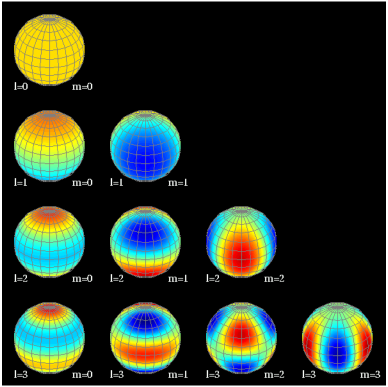
Irreps of compact groups

Theorem. Let $\mathcal{R}$ be a complete set of irreps of a compact group G. Then

1. Each $\rho \in \mathcal{R}$ is finite dimensional (the dimensionality is denoted d_ρ).
2. Each $\rho \in \mathcal{R}$ can be chosen to be unitary.
3. $\mathcal{R}$ is countable $\rightarrow$ we can talk about $\rho_1, \rho_2, \ldots$.
4. If $\mathcal{R}'$ is an alternative complete set of irreps of G, then there is a bijection $\gamma\colon \mathcal{R} \rightarrow \mathcal{R}'$ such that $\rho \cong \gamma(\rho)$.

The irreps of a compact group are essentially uniquely defined.

Example: The rotation group $SO(3)$



The irreps are given by the Wigner D matrices

$$D^{\ell}_{m,m'} = (-1)^m \sqrt{\frac{4\pi}{2\ell+1}}\, Y^m_\ell(\phi,\theta) e^{im'\psi}, \qquad m, m' \in \{-\ell, \dots, \ell\}\,.$$

The regular representation

Any group acts on itself by $g\colon x \mapsto gx$ and the corresponding representation

$$\mu_{\text{reg}} : f \mapsto f' \qquad f'(x) = f(g^{-1}x) \qquad f \in L_2(G)$$

called the **regular representation** of G.

Theorem. If G is compact, then

$$\mu_{\text{reg}} = \bigoplus_{\rho \in \mathcal{R}} \rho^{\oplus d_\rho}.$$

Invariants

General setting

1. We have a symmetry group G acting on a set $\mathcal{X}$.
2. The action extends to the space of functions, V.
3. We want to find invariants $\Upsilon : V \to \mathbb{R}$ to the action of G.

We assume that G is compact and V is a Hilbert space.

General setting

1. Consider the “translation operators”

$$T_g : f \mapsto f' \qquad f'(x) = f(g^{-1}(x)).$$

   These form a representation μ of G.

2. By complete reducibility,

$$\mu = \bigoplus_i \rho_i^{\oplus \kappa_\mu(\rho_i)},$$

   and we have a corresponding orthogonal decomposition of V

$$V = V_1 \oplus V_2 \oplus \ldots$$

$$V_i = W_{i,1} \oplus W_{i,2} \oplus \ldots \oplus W_{i,\kappa_\mu(\rho_i)}$$

   into subspaces that are invariant under the T_g action of G.

General setting

1. In any given $W_{i,j}$ subspace the action of G is

$$h \mapsto \rho_i(g)\, h.$$

2. Because ρ_i is unitary, setting $h = f{\downarrow}_{W_{i,j}}$

$$\Upsilon_i[f] := \|\rho_i(g)(h)\|^2 = \|h\|^2 \,,$$

so Υ_i is an invariant!

$\rightarrow$ We have as many invariants now as irreps in the decomposition. Actually, can also consider products of the form $f{\downarrow}^{*}_{W_{i,j_1}} \cdot f{\downarrow}_{W_{i,j_2}}$ (same i). Is this enough? How do we find out how V decomposes without all the abstract representation theory?

Example: $\mathbb{R}$ acting on $\mathbb{R}$

1. The action is

$$T_g : f \mapsto f' \qquad f'(x) = f(x-g) \qquad g \in \mathbb{R}.$$

2. The invariant subspaces are the 1D spaces

$$W_\omega = \operatorname{span}\left\{e^{-2\pi i\omega x}\right\}.$$

3. The projection of f to W_ω is the scalar

$$h_\omega = \int e^{2\pi i\omega x} f(x)\, dx.$$

4. The corresponding invariant is $\Upsilon_\omega = \|h_\omega\|^2$.

This looks suspiciously like the power spectrum.

$\mathrm{SO}(3)$ acting on S^2

1. Set $V = L_2(S^2)$ and

$$T_R : f \mapsto f' \qquad f'(x) = f(R^{-1}x) \qquad R \in \mathrm{SO}(3).$$

2. The invariant subspaces are

$$W_\ell = \mathrm{span}\left\{Y_\ell^m\right\}.$$

3. The projection of f to W_ℓ is $h \in \mathbb{C}^{2\ell+1}$ with components

$$h_\ell = \int\int Y_\ell^m(\theta,\varphi))^* f(\theta,\varphi)\, d\Omega(\theta,\varphi).$$

4. The corresponding invariant is $\Upsilon_\omega = h_\ell^\dagger h_\ell = \|h_\ell\|^2$.

This is just the spherical power spectrum.

Fourier transforms

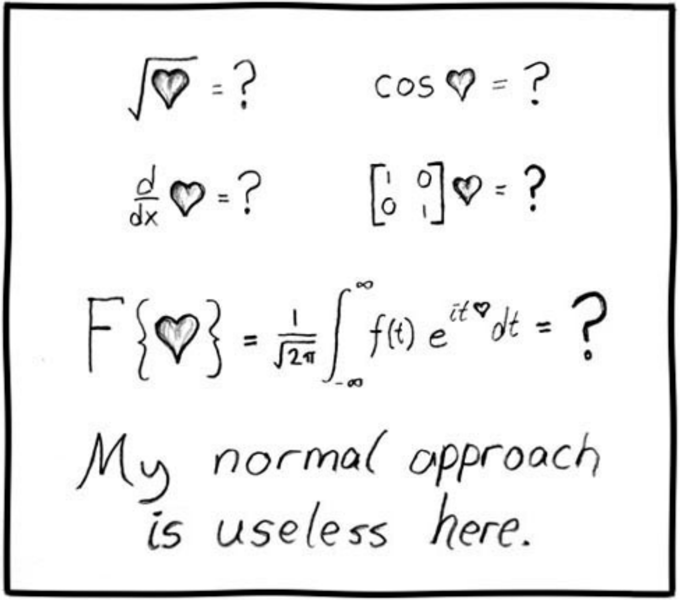
Fourier transform on $\mathbb{R}$

The Fourier transform of $f : \mathbb{R} \to \mathbb{C}$ is

$$\widehat{f}(\omega) = \int e^{-2\pi i \omega x} f(x)\, dx.$$

We have seen that $\left\{e^{-2\pi i \omega x}\right\}$ are exactly the irreps of $\mathbb{R}$, and $\widehat{f}(\omega)$ is the projection onto the W_ω invariant subspace.

How does this generalize?


My normal approach is useless here.

Fourier transform on G

Given a compact group G, and $f \in L_2(G)$, purely heuristically, define

$$\widehat{f}(\rho) = \int_{x \in G} \rho(x)\, f(x)\, d\mu(x) \qquad\qquad \rho \in \mathcal{R}.$$

This is weird because the Fourier components are now **matrices**.

Translation theorem

Theorem. Let $f \in L_2(G)$ and for some $g \in G$

$$f'(x) = f(g^{-1}x) = (T_g f)(x).$$

Then for any $\rho \in \mathcal{R}$

$$\widehat{f'}(\rho) = \int_{x \in G} \rho(g)\,\rho(x)\,f(x)\,d\mu(x) = \rho(g) \cdot \widehat{f}(\rho).$$

Translation theorem

Corollary. $\widehat{f}(\rho_i)$ is the projection of f to V_i, and its j'th column, $[\widehat{f}(\rho_i)]_j$ is the projection of f to $W_{i,j}$.

Corollary. The Fourier transform decomposes $L_2(G)$ into irreducible T_g invariant subspaces.

Convolution theorem

Theorem. Given $f, h \in L_2(G)$, define their **convolution** as

$$(f * h)(x) = \int f(xy^{-1})\, h(y)\, d\mu(y).$$

Then for any $\rho \in \mathcal{R}$

$$\widehat{f * h}(\rho) = \widehat{f}(\rho) \cdot \widehat{h}(\rho).$$

Correlation theorem

Theorem. Given $f, h \in L_2(G)$, define their **correlation** as

$$(f \star h)(x) = \int f(xy)\, h(y)^* \, d\mu(y).$$

Then for any $\rho \in \mathcal{R}$

$$\widehat{f \star h}(\rho) = \widehat{f}(\rho) \cdot \widehat{h}(\rho)^\dagger.$$

Back to invariants

Noncommutative power spectrum

The **power spectrum** of $f \in L_2(G)$ is

$$\widehat{a}(\rho) = \widehat{f}(\rho)^\dagger \cdot \widehat{f}(\rho).$$

Clearly invariant because

$$\widehat{f^\tau}(\rho)^\dagger \cdot \widehat{f^\tau}(\rho) = (\rho_\rho(t) \cdot \widehat{f}(\rho))^\dagger (\rho_\rho(t) \cdot \widehat{f}(\rho)) = \widehat{f}(\rho)^\dagger \cdot \widehat{f}(\rho).$$

The power spectrum is the FT of the (flipped) autocorrelation function

$$a(h) = \sum_{g \in G} f(gh^{-1}) f(g).$$

Exactly the same as invariants formed from the $f\downarrow_{W_{i,j}}$ on slide 36.

The noncommutative bispectrum

Recall the Clebsch-Gordan decomposition

$$\rho_1(\sigma) \otimes \rho_2(\sigma) = C_{\rho_1,\rho_2}\Big[\bigoplus_{\rho \in R_{\rho_1,\rho_2}} \bigoplus_{i=1}^{c(\rho_1,\rho_2,\rho)} \rho(\sigma)\Big] C^{\dagger}_{\rho_1,\rho_2}.$$

The bispectrum:

$$\widehat{b}_f(\rho_1,\rho_2) = C^{\dagger}_{\rho_1,\rho_2}\left[\widehat{f}(\rho_1) \otimes \widehat{f}(\rho_2)\right]^{\dagger} C_{\rho_1,\rho_2} \bigoplus_{\rho \in \Lambda_{\rho_1,\rho_2}} \bigoplus_{i=1}^{c(\rho_1,\rho_2,\rho)} \widehat{f}(\rho)$$

The bispectrum is the FT of the **triple correlation**

$$b(h_1,h_2) = \sum_{g \in G} f(gh_1^{-1}) f(gh_2^{-1}) f(g).$$

Completeness result

Theorem [Kakarala, 1992]. Let f and f' be a pair of complex valued integrable functions on a compact group G. Assume that $\widehat{f}(\rho)$ is invertible for each $\rho \in \mathcal{R}$. Then $f' = f^z$ for some $z \in G$ if and only if $b_f(\rho_1, \rho_2) = b_{f'}(\rho_1, \rho_2)$ for all $\rho_1, \rho_2 \in \mathcal{R}$.

- Generalizes to any Tatsuuma duality group (e.g., $\mathrm{ISO}(n)$)

The skew spectrum

The **skew spectrum** of $f : \mathbb{S}_n \to \mathbb{C}$ is the collection of matrices

$$\widehat{q}_h(\rho) = \widehat{r}_h^{\dagger}(\rho) \cdot \widehat{f}(\rho), \qquad \rho \in \mathcal{R}_G, \quad \widehat{\in} G,$$

with $r_h(g) = f(gh)\, f(g)$.

Unitarily equivalent to the bispectrum, but sometimes easier to compute [K., 2007]

Conclusions

Conclusions

Noncommutative harmonic analysis provides a canonical way to construct invariants to the action of compact groups on their homogeneous spaces.

Outstanding issues:

- What are the algebraic relationships between the components of the bispectrum?
- Can we prove completeness on homogeneous spaces?
- Can we extend the theory to noncompact groups?
- What are the smoothness properties of the bispectrum?
- How do we construct wavelets on groups?

A two-year postdoc position in available in my group at UChicago starting immediately.