Group theoretical methods in Machine Learning

Risi Kondor
Columbia University

Tutorial at ICML 2007

Tiger, tiger, burning bright
In the forests of the night,
What immortal hand or eye
Dare frame thy fearful symmetry?

William Blake
(1757-1827)

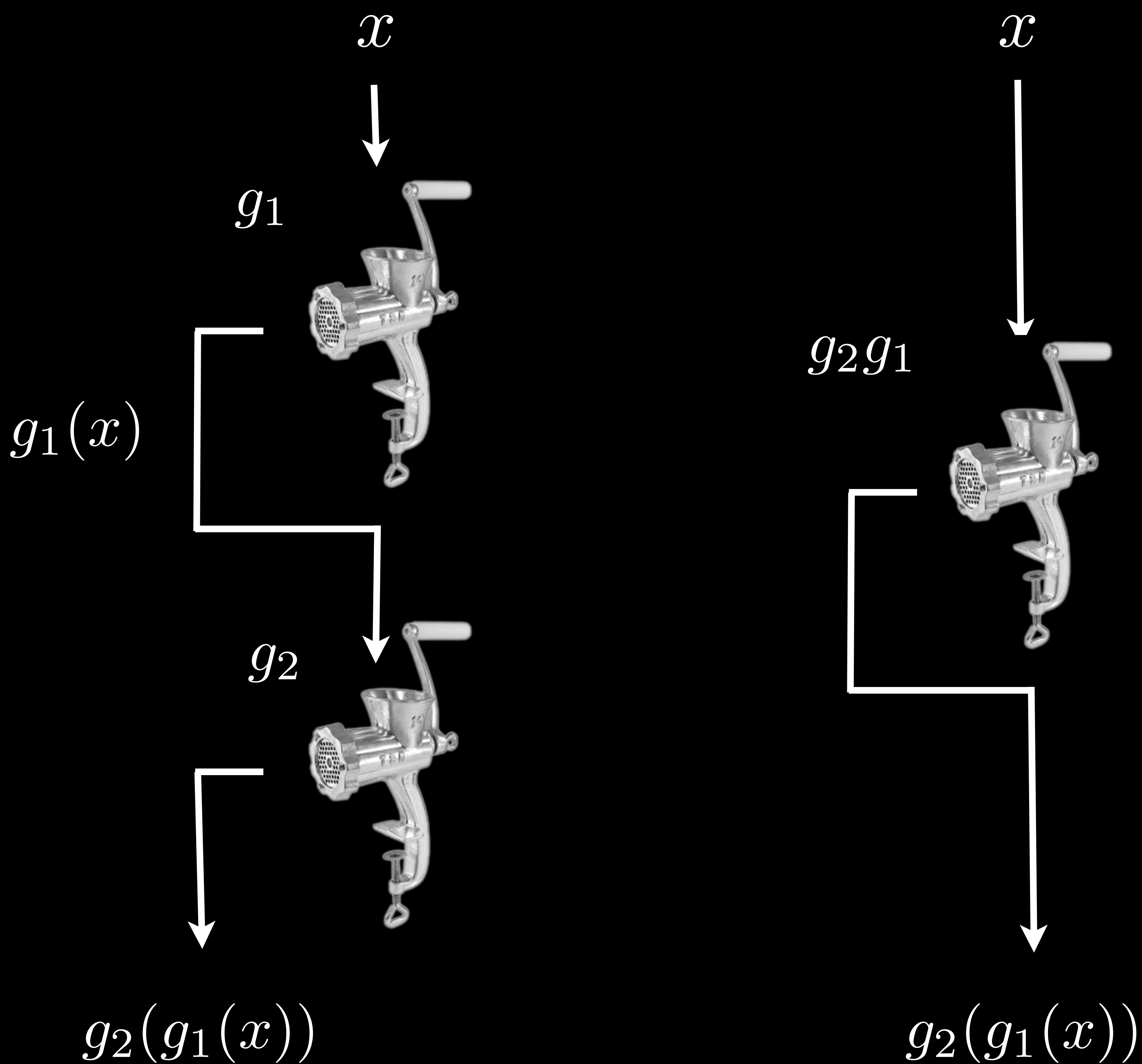
Evariste Galois
(1811-1827)

Niels Henrik Abel
(1802-1829)

$g\colon \mathcal{X} \to \mathcal{X}$

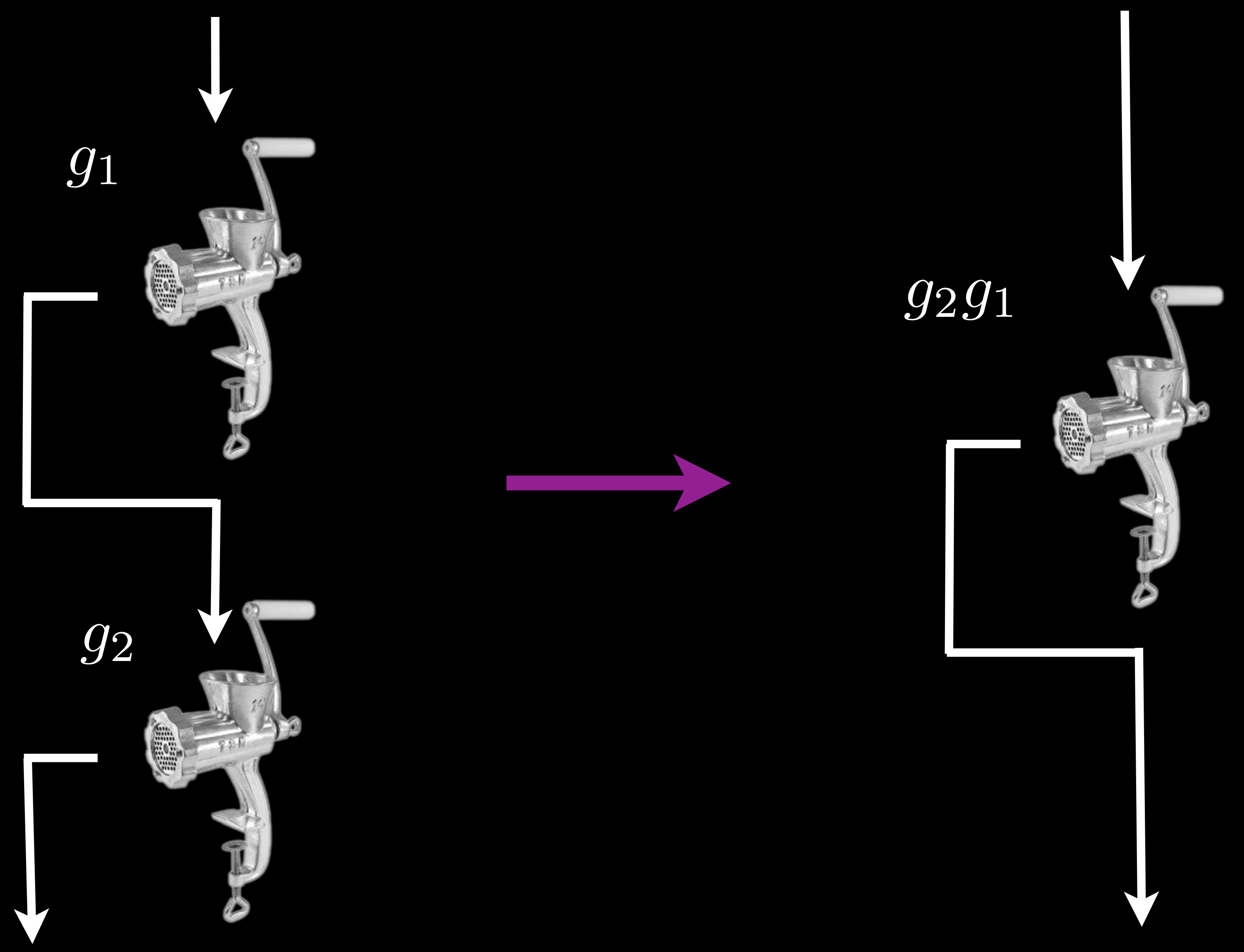
x

$g(x)$



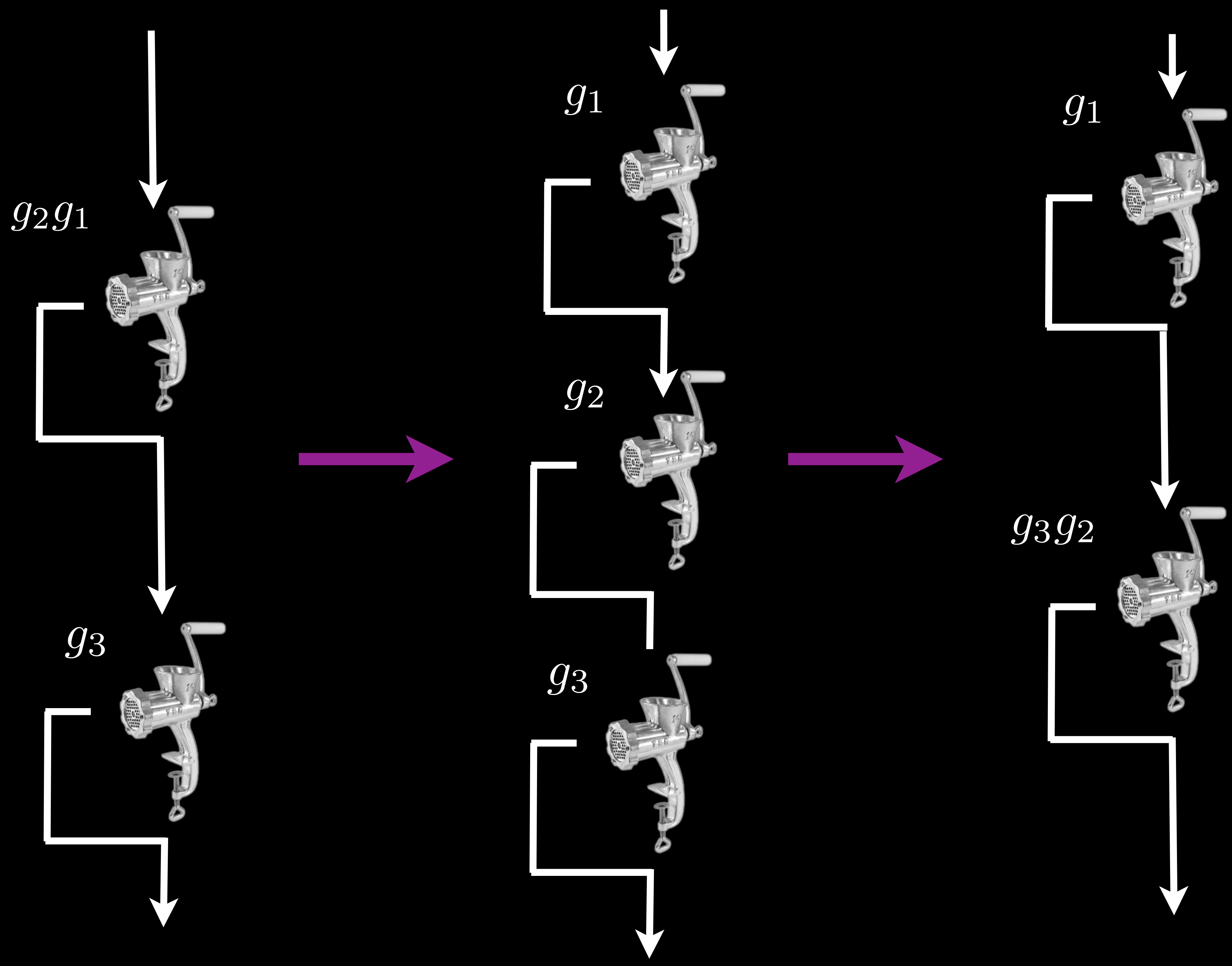
x
g_1
$g_1(x)$
g_2
$g_2(g_1(x))$
x
$g_2 g_1$
$g_2(g_1(x))$

1.



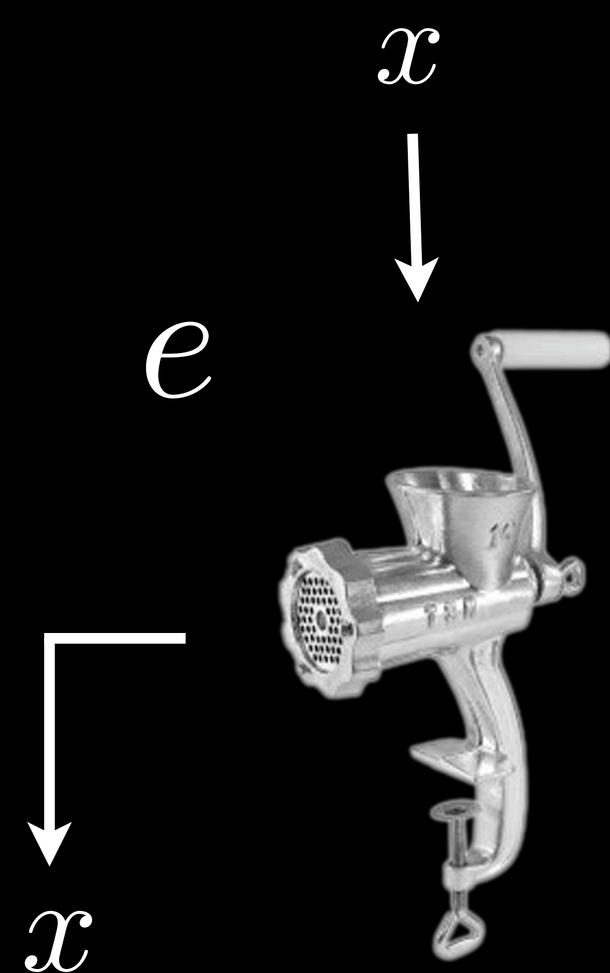
For any $g_1, g_2 \in G$, $g_2 g_1 \in G$.

2.



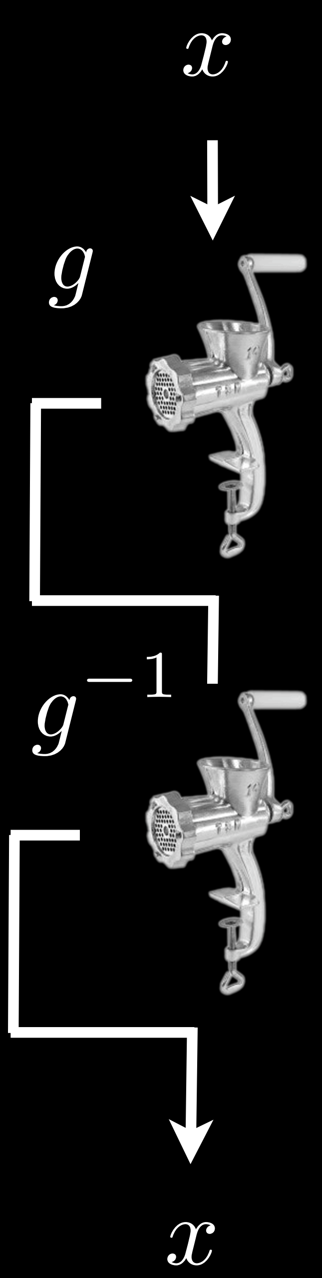
For any $g_1, g_2, g_3 \in G$, $g_3(g_2 g_1) = (g_3 g_2) g_1$.

3.



There exists an identity $e \in G$ such that $eg = ge = g$ for any $g \in G$.

4.



For any $g \in G$ there is a $g^{-1} \in G$ such that $g^{-1}g = e$.

A set G endowed with multiplication is a group if

1. $g_1 g_2 \in G$ (closure);
2. $g_3(g_2 g_1) = (g_3 g_2) g_1$ (associativity);
3. $e \in G$ such that $eg = ge = g$ (identity);
4. there exists a $g^{-1} \in G$ such that $g^{-1} g = g\, g^{-1} = e$ (inverses).

G is a commutative (Abelian) group if $g_1\, g_2 = g_2\, g_1$

Example 1

The cyclic group $\mathbb{Z}_n = \{0, 1, 2, \ldots, n-1\}$

$$xy = x + y \mod n$$

Example 2

Klein's Viergruppe $V = \{1, i, j, k\}$

	1	i	j	k
1	1	i	j	k
i	i	1	k	j
j	j	k	1	i
k	k	j	i	1

$$V \cong \mathbb{Z}_2 \times \mathbb{Z}_2$$

Example 3

The quaternion group $Q = \{1, i, j, k, -1, -i, -j, -k\}$

	1	i	j	k	-1	$-i$	$-j$	$-k$
1	1	i	j	k	-1	$-i$	$-j$	$-k$
i	i	-1	k	$-j$	$-i$	1	$-k$	j
j	j	$-k$	-1	i	$-j$	k	1	$-i$
k	k	j	$-i$	-1	$-k$	$-j$	i	1
-1	-1	$-i$	$-j$	$-k$	1	i	j	k
$-i$	$-i$	1	$-k$	j	i	-1	k	$-j$
$-j$	$-j$	k	1	$-i$	j	$-k$	-1	i
$-k$	$-k$	$-j$	i	1	k	j	$-i$	-1

$$-1^2 = 1$$

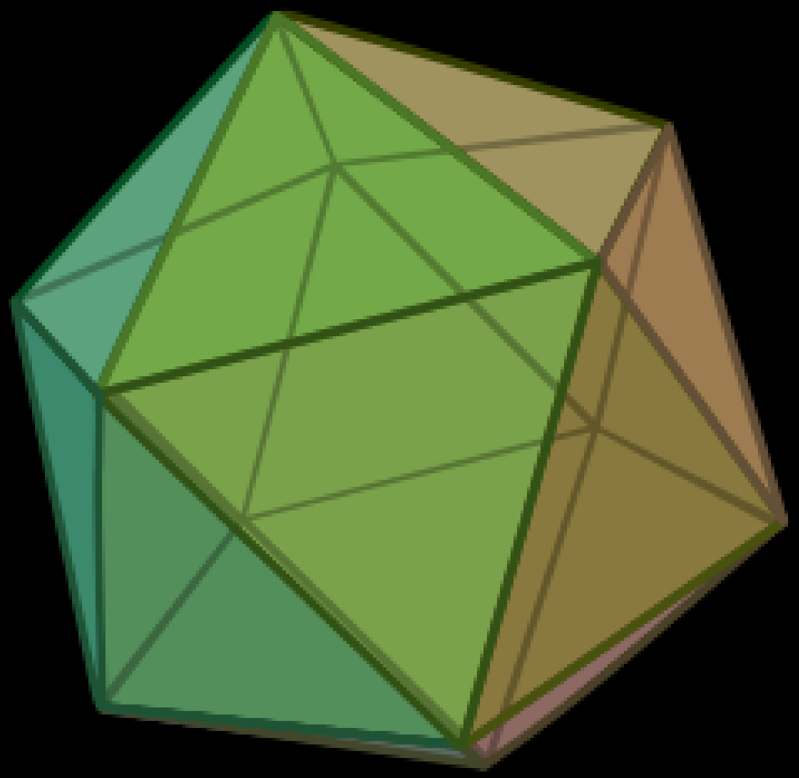
$$(-1)a = a(-1) = -a$$

$$i^2 = j^2 = k^2 = -1$$

$$ij = k$$

Example 4

The icosahedron group $I_h \cong A_5$



Example 5

The symmetric groups $\mathbb{S}_n$

group of bijections

$$\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$$

i.e., permutations of n objects

Example 6

The integers $\mathbb{Z}$

$$xy = x + y$$

Example 7

The reals $\mathbb{R}$ and the Euclidean vector spaces $\mathbb{R}^n$

$$xy = x + y$$

Example 8

The rotation groups $\mathrm{SO}(n)$

group of $n \times n$ orthogonal matrices of det 1

Example 9

The Euclidean group $\mathrm{ISO}(n)$ and group of rigid body motions $\mathrm{ISO}^+(n)$

Erlangen program (1872):

"geometry is the study of properties invariant under a group"

Example 10

The special unitary groups $\mathrm{SU}(n)$

group of $n \times n$ unitary matrices of determinant 1

Example 11

The general linear group $\mathrm{GL}(n)$

group of $n \times n$ invertible matrices

Example 12

$$G = (\mathbb{Z}_3^7 \times \mathbb{Z}_2^{11}) \rtimes ((A_8 \times A_{12}) \rtimes \mathbb{Z}_2)$$

Example 13

The Monster group M

$$|M| = 808017424794512875886459904961 7\ldots$$
$$\ldots 1075700575436800000000 0$$

What immortal hand or eye
Dare frame thy fearful symmetry?

Finite groups $\mathbb{Z}_n \; V \; Q \; M \; \mathbb{S}_n$

Infinite groups

Countable groups $\mathbb{Z}$

Continuous groups

Lie groups

compact $\mathrm{SO}(n) \; \mathrm{SU}(n)$

non-compact $\mathbb{R}^n \; \mathrm{ISO}^+(n)$

Closed fields

Fields

Rings

Groups → commutative / non-commutative

Semigroups

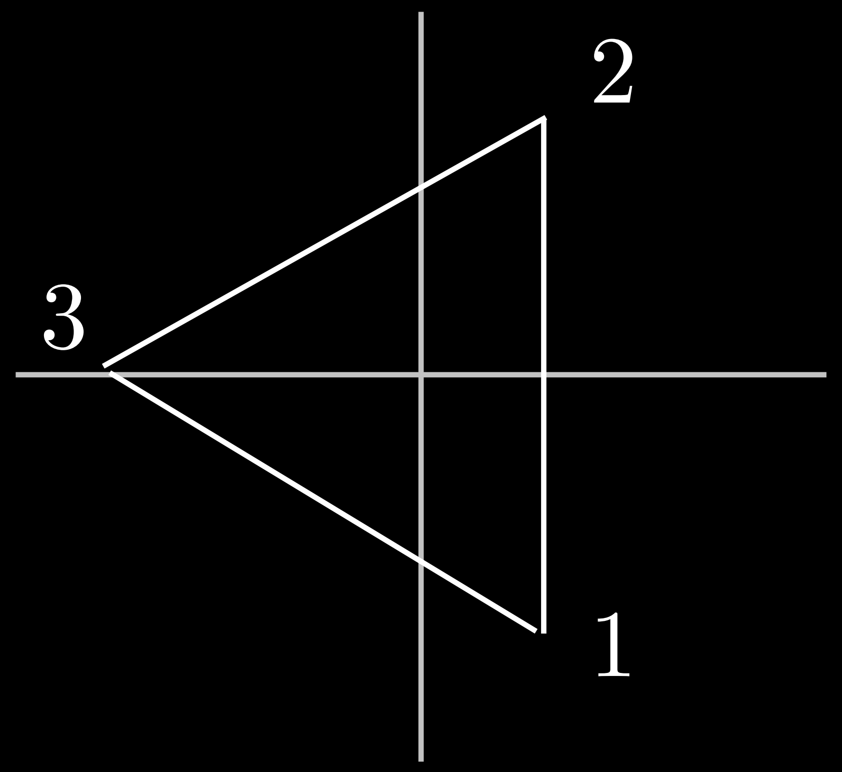
Representations

The idea is to “model” groups by $\rho\colon G \to \mathbb{C}^{d\times d}$ such that

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

Example

$$\mathbb{S}_3$$



$$\rho(e) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$\rho((12)) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

$$\rho((123)) = \begin{pmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{pmatrix}$$

Example

A representation of $Q = \{1, i, j, k, -1, -i, -j, -k\}$:

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

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

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

$$\rho(k) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

$$\rho(-x) = -\rho(x)$$

$$\rho(i)^2 = \rho(j)^2 = \rho(k)^2 = \rho(-1)$$

$$\rho(-1)\,\rho(x) = \rho(-x)$$

$$\rho(i)\,\rho(j) = \rho(k)$$

What are the “fundamental” representations?

1. Two representations are equivalent if

$$\rho_1(x) = T^{-1}\rho_2(x)\,T \qquad \forall\, x \in G$$

2. A representation is reducible if

$$T^{-1}\rho(x)\,T = \begin{pmatrix} \rho_1(x) & 0 \\ 0 & \rho_2(x) \end{pmatrix} \qquad \forall\, x \in G$$

A complete set of inequivalent irreducible representations we denote $\mathcal{R}$.

For any compact group the irreducibles $\rho \in \mathcal{R}$ can always be chosen to be unitary, $\rho(x)^{-1} = \rho(x)^{\dagger}$.

Maschke/Wedderburn theorem

Any representation ρ of a compact group reduces essentially uniquely into a direct sum

$$\rho(x) = T^{-1}\left[\bigoplus_{\rho'} \rho'(x)\right] T$$

of irreducible representations.

The Clebsch-Gordan decomposition

$$\rho_1(x) \otimes \rho_2(x) = C^{-1}_{\rho_1,\rho_2} \left[\bigoplus_{\rho'} \rho'(x) \right] C_{\rho_1,\rho_2}$$

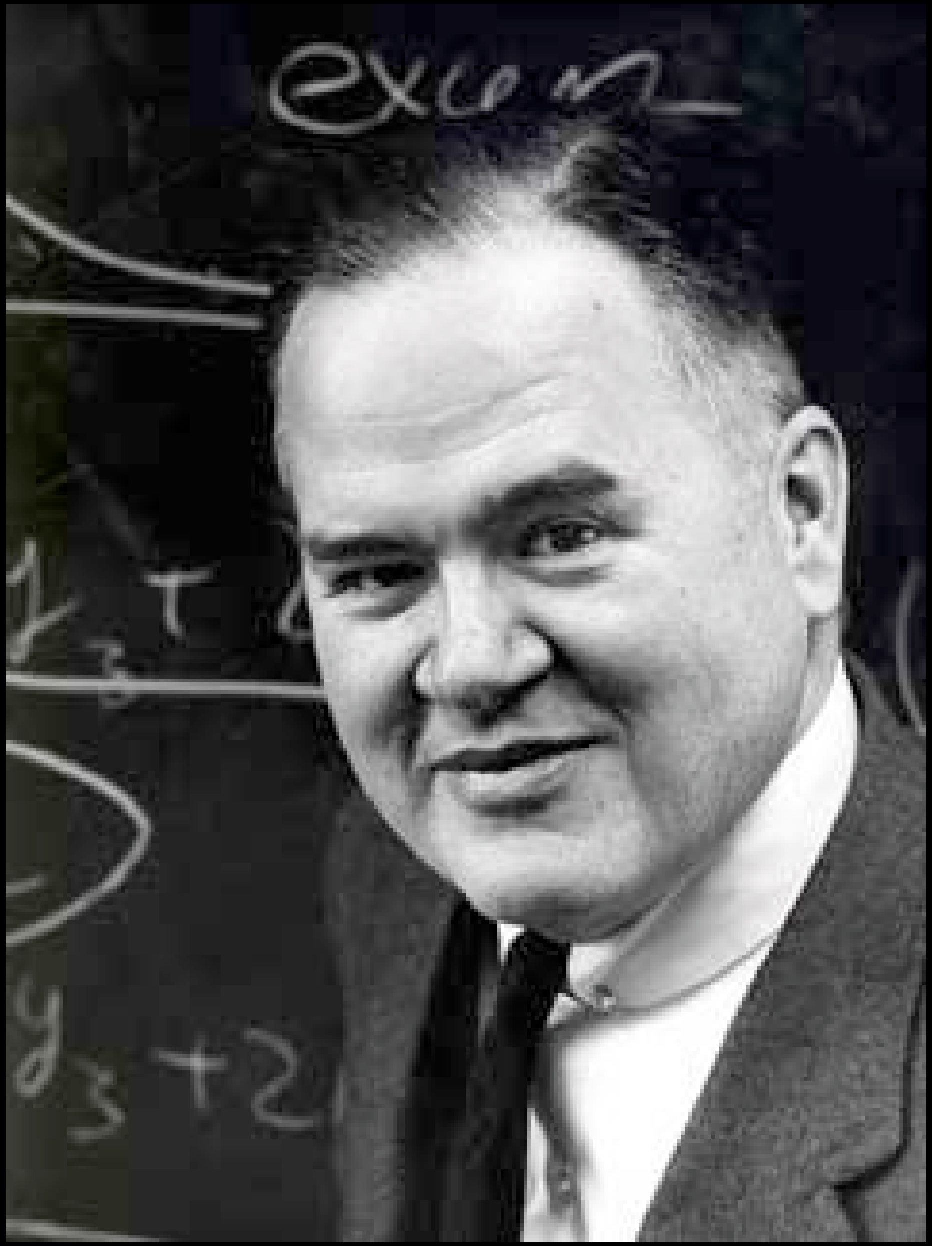
Harmonic analysis

$$f(x) = \sum_{k=-\infty}^{\infty} f_k \, e^{ikx}$$

$$f_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-ikx} f(x) \, dx$$

$$\widehat{f}(k) = \int e^{-i2\pi kx} f(x) \, dx$$

Joseph Fourier
1768-1830



John Wilder Tukey
1915-2000

"An algorithm for the machine calculation of complex Fourier series,", James W. Cooley and John W. Tukey, *Math. Comput.* **19**, 297–301 (1965)

Let $\mathbb{C}G$ be the space of functions $f : G \to \mathbb{C}$.

The left-translate of f is defined $f^z(x) = f(z^{-1}x)$.

Are there any subspaces of $\mathbb{C}G$ which are invariant under $f \mapsto f^z$?

Translation is a linear operator, so in vector form it acts by

$$\boldsymbol{v} \mapsto \rho(z)\,\boldsymbol{v}.$$

In particular,

$$\boldsymbol{f}^z = \rho_r(z)\,\boldsymbol{f},$$

where

$$[\rho_r(z)]_{y,x} = \begin{cases} 1 & \text{if} \;\; zx = y \\ 0 & \text{otherwise} \end{cases}$$

is the regular representation.

If we define

$$\widehat{f}(\rho) = \sum_{x \in G} f(x)\, \rho(x) \qquad\qquad \rho \in \mathcal{R}$$

then

$$\widehat{f^z}(\rho) = \rho(z)\, f(\rho)$$

So the columns of $\widehat{f}(\rho)$ are projections onto minimal left-translation invariant subspaces.

The Fourier transform on a finite group is

$$\widehat{f}(\rho) = \sum_{x \in G} f(x)\, \rho(x) \qquad\qquad \rho \in \mathcal{R}$$

Note similarity to

$$\widehat{f}(k) = \int e^{-i2\pi kx} f(x)\, dx$$

The Fourier transform on a group is

$$\widehat{f}(\rho) = \sum_{x \in G} f(x)\, \rho(x) \qquad\qquad \rho \in \mathcal{R}$$

The inverse transform is

$$f(x) = \frac{1}{|G|} \sum_{\rho \in \mathcal{R}} d_\rho \operatorname{tr}\left[\widehat{f}(\rho)\, \rho(x^{-1})\right]$$

$$\widehat{f}(\rho) = \sum_{x \in G} f(x)\, \rho(x) \qquad f(x) = \frac{1}{|G|} \sum_{\rho \in \mathcal{R}} d_\rho \operatorname{tr}\left[\widehat{f}(\rho)\, \rho(x^{-1})\right]$$

1. Linearity: $\widehat{f+g} = \widehat{f} + \widehat{g}$
2. Unitarity: $\langle f, g \rangle = \langle \widehat{f}, \widehat{g} \rangle$
3. Left-translation: $\widehat{f^z(\rho)} = \rho(z)\, \widehat{f}(\rho)$
4. Convolution: $\widehat{f * g}(\rho) = \widehat{f}(\rho)\, \widehat{g}(\rho)$

How about right translation $f^{(z)}(x) = f(xz^{-1})$?

$\widehat{f}^{(z)}(\rho) = \widehat{f}(\rho)\,\rho(z)$ so right-translation preserves the subspaces spanned by the rows of $\widehat{f}(\rho)$.

The isotypal decomposition is

$$\mathbb{C}G = \bigoplus_{\rho \in \mathcal{R}} V_\rho$$

where the V_ρ are both left- and right-invariant.

This decomposition is independent of the choice of $\mathcal{R}$!

The Fourier transform $\mathcal{F}\colon f \to \widehat{f}$ is an isomorphism

$$\mathcal{F}\colon \mathbb{C}G \to \bigoplus_{\rho \in \mathcal{R}} \mathbb{C}^{d_\rho \times d_\rho}.$$

The Laplacian can quantify the complexity of functions:

$$\Delta = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \ldots + \frac{\partial^2}{\partial x_D^2}$$

$$\Delta\, e^{-ikx} = -k^2\, e^{-ikx}$$

$$[\Delta]_{i,j} = \begin{cases} 1 & \text{if} \quad i \sim j \\ -\, d_i & \text{if} \quad i = j \\ 0 & \text{otherwise} \end{cases}$$

$$\boldsymbol{f}^\top \Delta\, \boldsymbol{f} = -\frac{1}{2} \sum_{i \sim j} (f(i) - f(j))^2$$

For example, in regularized risk minimzation

$$f = \arg\min \left[\sum_{i=1}^{m} L(f(x_i), y_i) + \langle f, f \rangle_{\mathcal{H}} \right]$$

where $\mathcal{H}$ is the RKHS induced by a kernel k.

My favorite kernel on graphs is the diffusion kernel $k(x, x') = \left[e^{\beta \Delta} \right]_{x,x'}$, giving rise to the regularizer

$$\langle f, f \rangle = \langle e^{-\beta\Delta/2} f,\ e^{-\beta\Delta/2} f \rangle$$

So what would the Laplacian on a group look like?

The Laplacian must be a self-adjoint, invariant operator:

$$\langle g, \Delta f \rangle = \langle \Delta g, f \rangle \qquad \langle g, f \rangle = \int_G (g(x))^* f(x)\, d\mu$$

$$\Delta f^z = (\Delta f)^z \qquad \Delta f^{(z)} = (\Delta f)^{(z)}$$

Theorem

If Δ is a left- and right-invariant operator then

$$\Delta = \bigoplus_{\rho \in \mathcal{R}} \alpha_\rho \, \mathbf{1}_\rho$$

In particular, $\Delta \widehat{f}(\rho) = \alpha_\rho \, \widehat{f}(\rho)$.

$\Delta f = \ell * f$ where

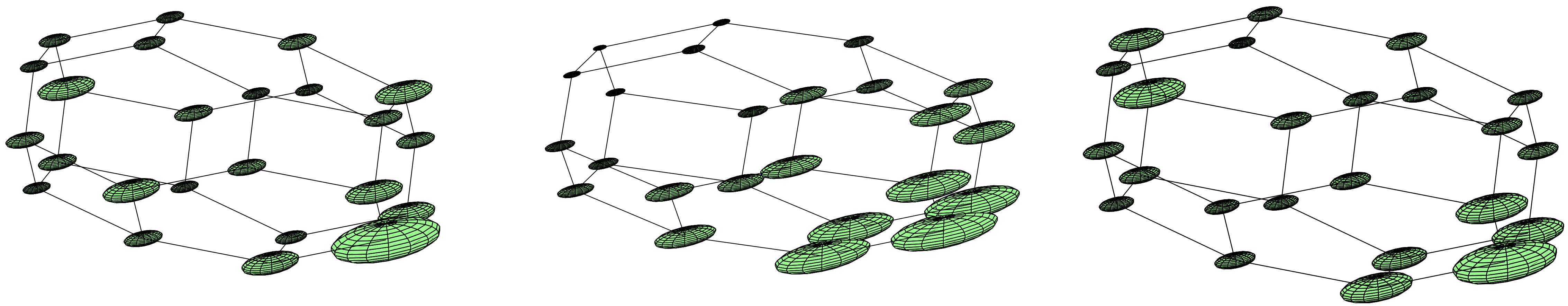
$$\widehat{\ell}(\rho) = \alpha_\rho \, I \quad \leftrightarrow \quad \ell^z = \ell^{(z)} \quad \leftrightarrow \quad \ell(z^{-1} x \, z) = \ell(x)$$

i.e., ℓ is a class function.

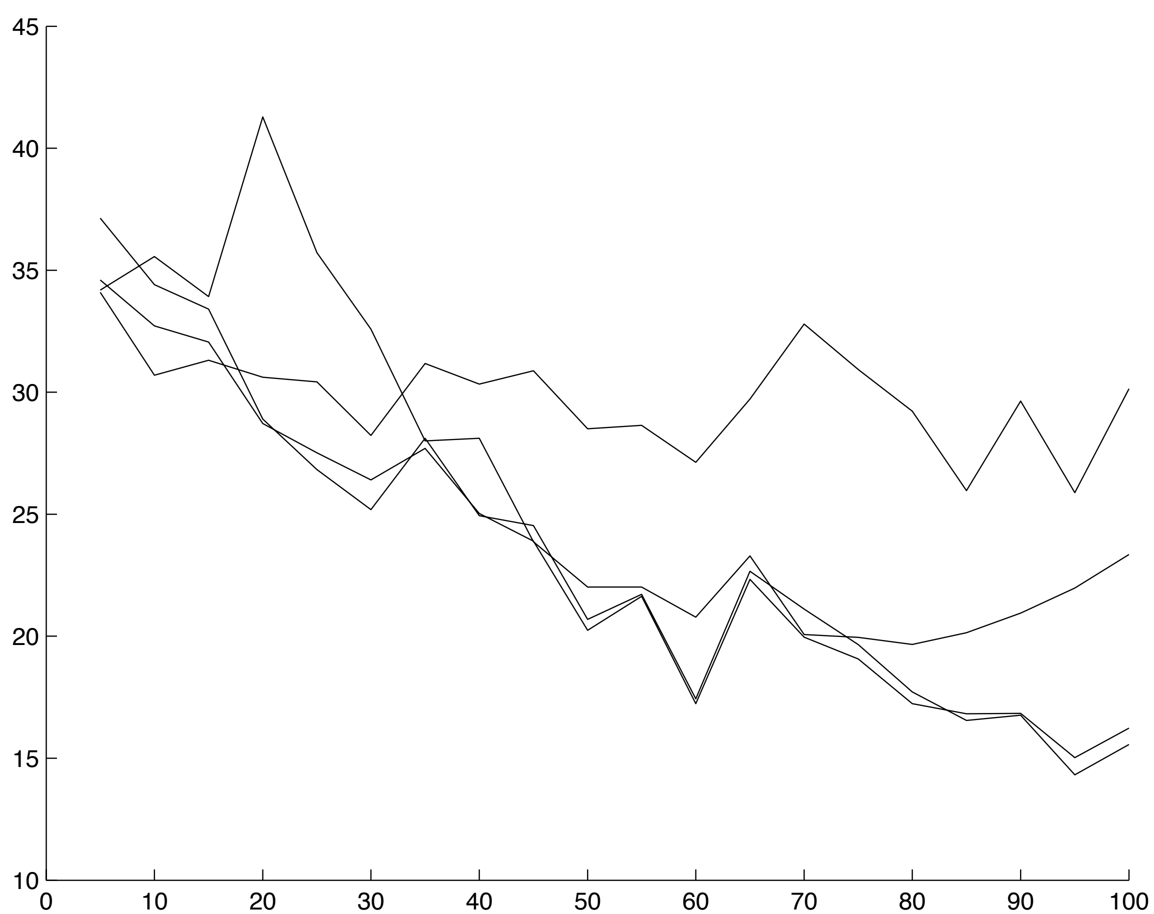
Example: data on $\mathbb{S}_n$

ranking	votes	ranking	votes	ranking	votes	ranking	votes	ranking	votes
54321	29	45321	31	35421	71	25431	35	15432	40
54312	67	45312	54	35412	61	25413	34	15423	35
54231	37	45231	34	35241	41	25341	40	15342	36
54213	24	45213	24	35214	27	25314	21	15324	17
54132	43	45132	38	35142	45	25143	106	15243	70
54123	28	45123	30	35124	36	25134	79	15234	50
53421	57	43521	91	34521	107	24531	63	14532	52
53412	49	43512	84	34512	133	24513	53	14523	48
53241	22	43251	30	34251	62	24351	44	14352	51
53214	22	43215	35	34215	28	24315	28	14325	24
53142	34	43152	38	34152	87	24153	162	14253	70
53124	26	43125	35	34125	35	24135	96	14235	45
52431	54	42531	58	32541	41	23541	45	13542	35
52413	44	42513	66	32514	64	23514	52	13524	28
52341	26	42351	24	32451	34	23451	53	13452	37
52314	24	42315	51	32415	75	23415	52	13425	35
52143	35	42153	52	32154	82	23154	186	13254	95
52134	50	42135	40	32145	74	23145	172	13245	102
51432	50	41532	50	31542	30	21543	36	12543	34
51423	46	41523	45	31524	34	21534	42	12534	35
51342	25	41352	31	31452	40	21453	24	12453	29
51324	19	41325	23	31425	42	21435	26	12435	27
51243	11	41253	22	31254	30	21354	30	12354	28
51234	29	41235	16	31245	34	21345	40	12345	30

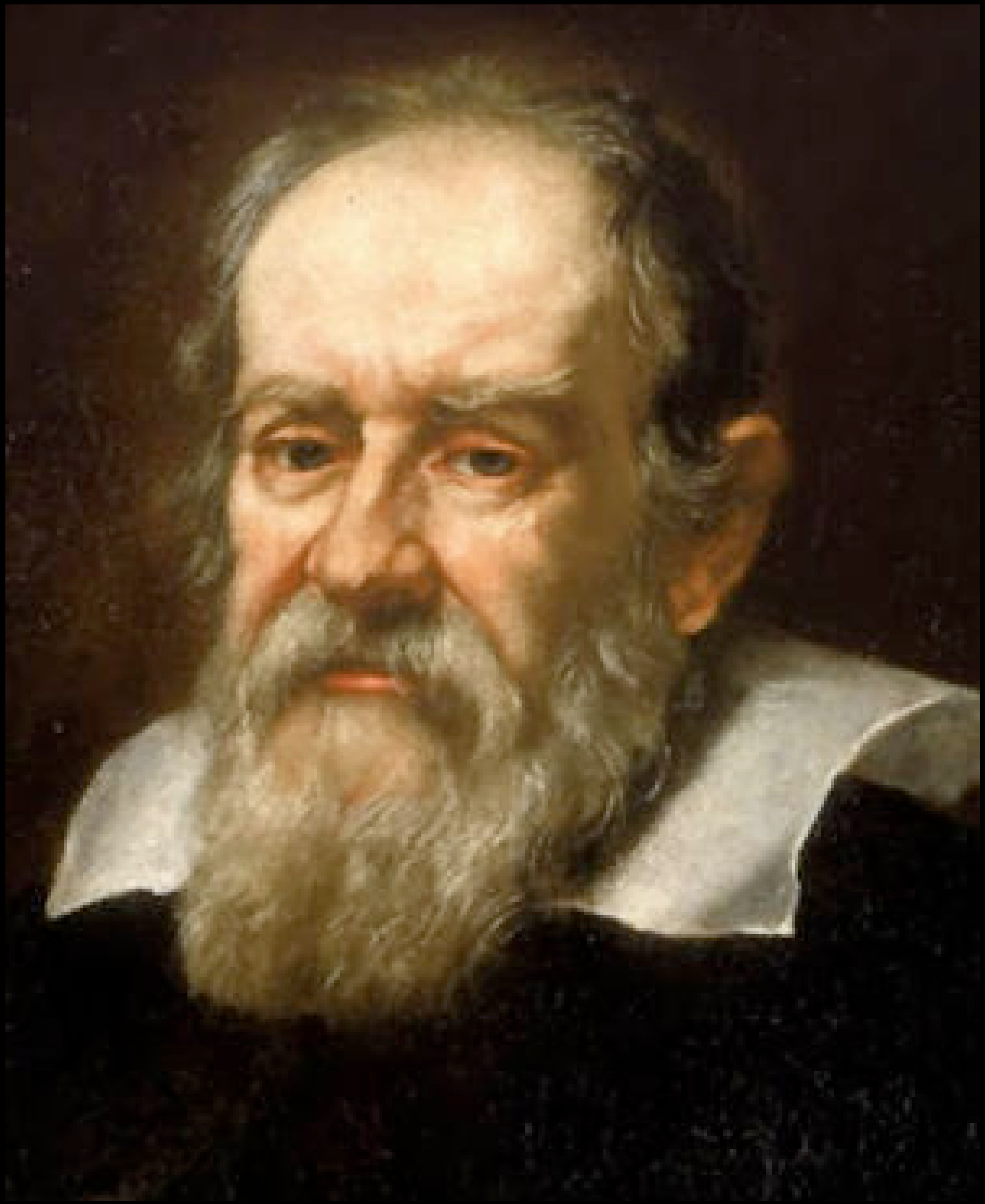
American Psychological Association voting data [Diaconis]



The kernels generated by transpositions, adjacent transpositions and cyclically adjacent transpositions.



Invariance in Physics



Galileo Galilei
(1564-1642)

The invariance group of classical Physics is

$$\mathrm{ISO}(3) \times \mathbb{R}$$

To preserve

$$(\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2 - c^2(\Delta t^2)$$

relativity adopted the Lorentz group $\mathrm{SO}^+(1,3)$

$$\begin{pmatrix} ct' \\ x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cosh(\beta) & -\sinh(\beta) & 0 & 0 \\ -\sinh(\beta) & \cosh(\beta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} ct \\ x \\ y \\ z \end{pmatrix}$$

$$\beta = \log\left(\frac{1 + v/c}{\sqrt{1 - v^2/c^2}}\right)$$

Albert Einstein
(1879-1955)

Symmetry implies conservation:

$$\frac{dJ}{dt} = -\frac{\partial \mathcal{H}}{\partial \theta}$$

(roughly)

time → energy

space → momentum

rotation → angular mom.

Emmy Noether
(1882-1935)

Eugene Wigner
(1902-1995)

symmetries	→	unitary op.
observables	→	generators
pure states	→	dimensions of irreps

Murray Gell-Man
1929-

Quantum Chromodynamics

$$\mathrm{SU}(3),\ \mathrm{SU}(6), \ldots$$

lead to quarks and even stranger animals...