Pairwise Symmetry Decomposition Method for Generalized Covariance Analysis

IBM Research, Tokyo Research Lab.
Tsuyoshi Idé （井手 剛）

Summary:
We generalize the notion of covariance using group theory.

covariance → **theoretical properties** → **generalization**

$$\langle (x-\bar{x})(y-\bar{y}) \rangle$$

where

$$\langle \bullet \rangle = \int \bullet \, p(x,y)dxdy$$

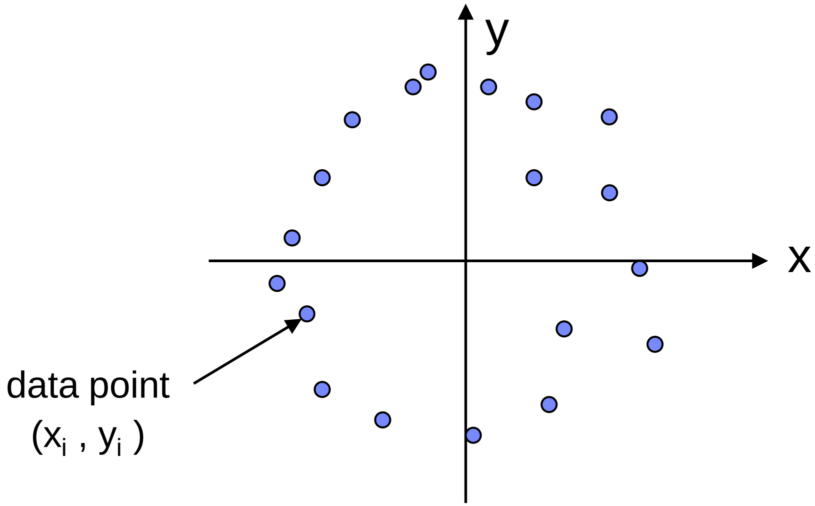
2-variate cross cumulant

An irreducible representation of a group

2-variate CC as irreducible representations

New approach to pattern recognition

Motivation.
The traditional covariance cannot capture nonlinearities.



- **The traditional covariance cannot capture nonlinearities.**

$$C_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

# of data points

  - C_{xy} would be *useless* in this case.

- **We wish to explicitly define useful metrics for nonlinear correlations.**
  - cf. kernel methods are black-boxes

The lowest order cross-cumulant (CC) is identical to the covariance. A generalized covariance could be higher-order CC.

Multivariate systems can be described with PDF.

$$p(\mathbf{x}) = p(x_1, x_2, \dots, x_n)$$

↓

The cumulants of p completely characterize p.

By definition,

$$\langle x_i x_j \rangle_{\mathrm{C}} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle$$

↓

2nd order CC is identical to the covariance

⇒

Notation of "cumulant average"

$$\langle {x_{i_1}}^{a_1} {x_{i_2}}^{a_2} \cdots {x_{i_k}}^{a_k} \rangle_{\mathrm{C}}$$

‖

$(a_1, \dots, a_k)$-th order cumulant w.r.t. $(x_1, \dots, x_k)$

principle #1

Use higher-order CC for generalizing the covariance

* We assume zero-mean data hereafter.

Relation between the covariance and symmetries.
The axioms of group can be used to characterize the symmetries.

The covariance can be viewed as an inner product of two vectors, $|p\rangle$ and $|xy\rangle$.

PDF

symmetric w.r.t. x and y, etc.

$|xy\rangle$ has potentially meaningful symmetries

Inner product?

$$\langle p| \cdot |xy\rangle = \langle p|xy\rangle \equiv \int d\boldsymbol{r} p(\boldsymbol{r}) xy.$$

A collection of such symmetry operations may be used for characterizing the symmetries.

What is the guiding principle to define it?

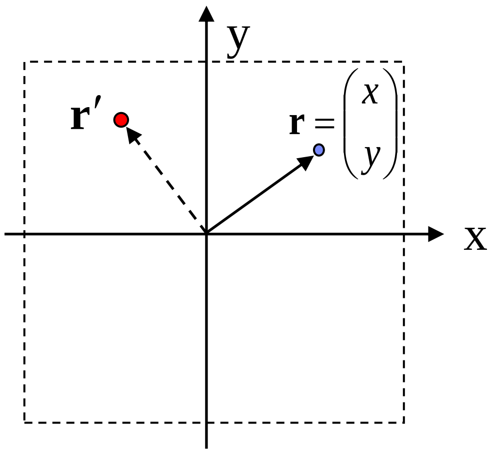
The axioms of group can be the guiding principle.

***Closure*, *Associativity*,**

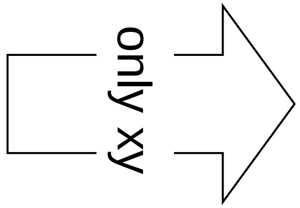
***Identity*, *Inverse*.**

The set of OPs is almost unique --- C_{4v} is the most appropriate group for characterizing the pairwise correlations.

- **Requirements on the group *G***
  - *G* should include OPs describing the symmetries within the *xy* plane.
  - G should include an OP to exchange *x* with *y*.



Point group is a natural choice



Most general one is a group named C_{4v}

Point group ?

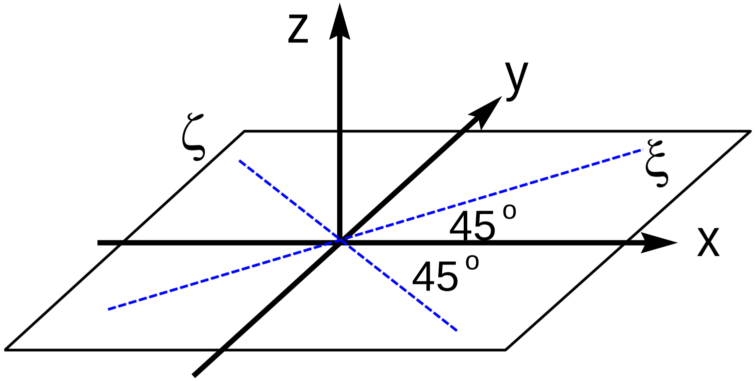
only rotations and mirror reflections

What is the C_{4v} group ?
It contains 8 symmetry operations within the xy space.

- C_4: 90^o rotation around the z axis
- ${C_4}^3$: 270^o rotation around the z axis
- C_2: 180^o rotation around the z axis

- σ_x: mirror reflection about the zx plane
- σ_y: mirror reflection about the zy plane
- σ_ξ: mirror reflection about the ξz plane
- σ_η: mirror reflection about the ηz plane

- I: identity operation (do nothing)



For $g \in C_{4v}$, $|f\rangle \in \mathcal{H}$,
the operation of g on $|f\rangle$
is defined by

$$g|f\rangle = \int d\boldsymbol{r}\ g|\boldsymbol{r}\rangle f(\boldsymbol{r}) = \int d\boldsymbol{r}|\boldsymbol{r}\rangle f(g^{-1}\boldsymbol{r})$$

Only a single IRR has been used for recognizing correlation patterns. We unveil the other representations !

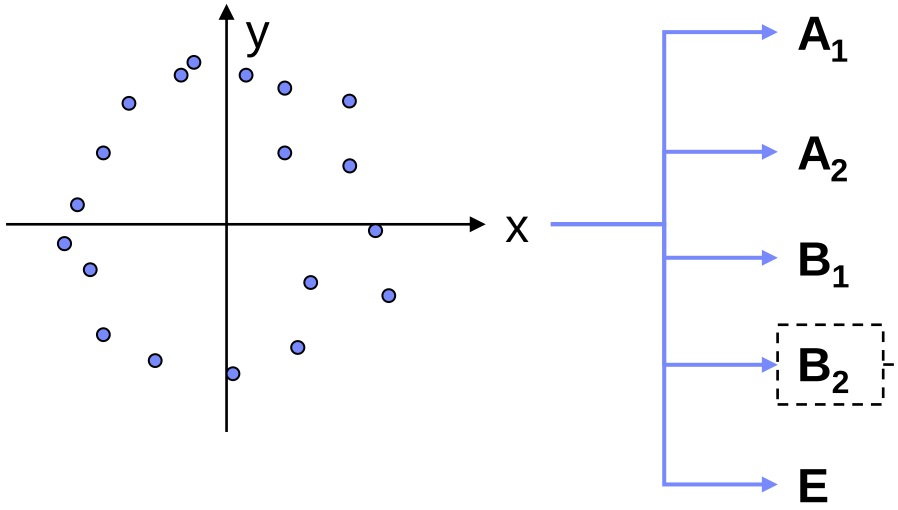
One can easily prove:

$|xy\rangle$ spans the B_2 irreducible representation (IRR) space of C_{4v}. Thus, $\langle p|xy\rangle$ is a B_2 IRR.

Further, a “symmetry decomposition theorem” holds:

any correlation pattern

linear combination of the *five* IRRs



Only the B_2 component has been used so far.

principle #2

Find the other IRRs, which haven’t been used so far.

The two consequences lead to the definition of the generalized covariances, which are symmetrized cross cumulants in the C_{4v} sense.

2-variate cross cumulants

$$\left\langle x^{\mu} y^{\nu} \right\rangle_{C}, \quad \mu + \nu \leq 4$$

principle #1

irreducible representations of C_{4v}

A_1, A_2, B_1, (B_2), E

principle #2

Construct IRRs as linear combinations of CCs

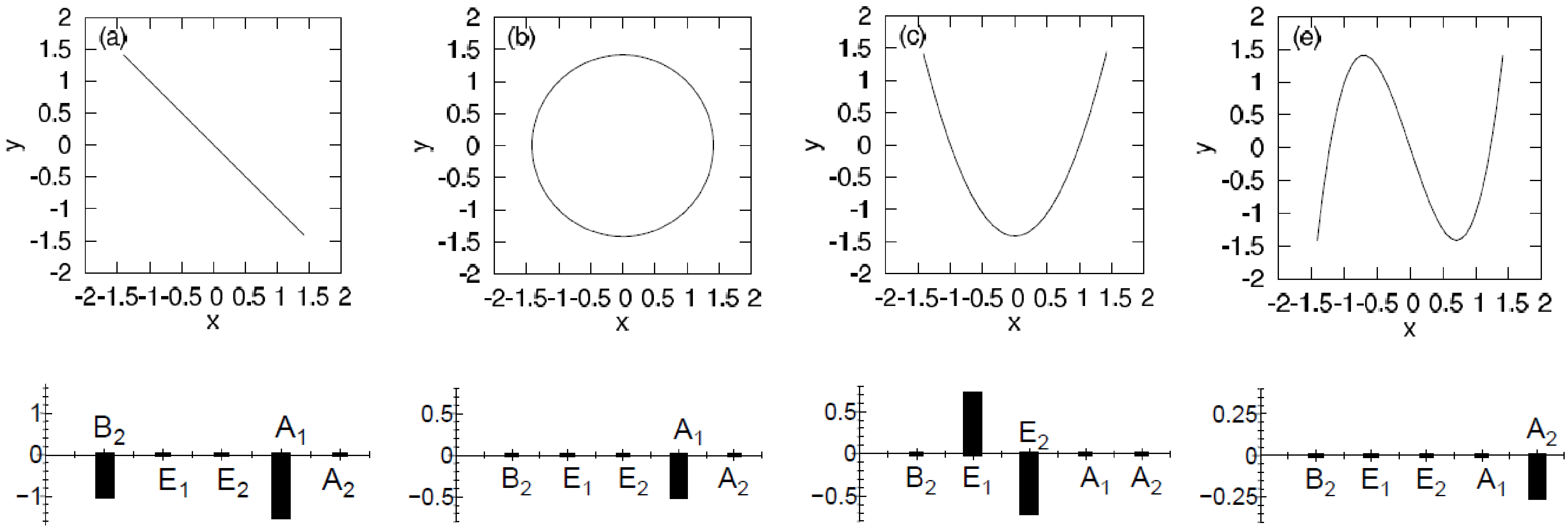
result:

$$
\begin{aligned}
\mathrm{C}(\mathrm{B}_2) &= \langle xy \rangle_c \\
\mathrm{C}(\mathrm{E}_1) &= \left[\langle xy^2 \rangle_c + \langle x^2 y \rangle_c \right] /2 \\
\mathrm{C}(\mathrm{E}_2) &= \left[\langle xy^2 \rangle_c - \langle x^2 y \rangle_c \right] /2 \\
\mathrm{C}(\mathrm{A}_1) &= \langle x^2 y^2 \rangle_c \\
\mathrm{C}(\mathrm{A}_2) &= \left[\langle xy^3 \rangle_c - \langle x^3 y \rangle_c \right] /2
\end{aligned}
$$

Note

- There is arbitrariness in prefactors
- x and y should be standardized (unit variance) to be scale invariant

Experiment with Lissajous' trajectories. The generalized covariances detect the nonlinearities while the standard covariance $C(B_2)$ fails.



- $C(B_2)$ should be minus 1 due to the perfect Inverse linear correlation

- $C(B_2)$ fails to capture the correlation
- $C(A_1)$ succeeds to detect the nonlinear correlation

- $C(B_2)$ fails to capture the correlation
- $C(E_2)$ and $C(E_2)$ succeed to detect the nonlinear correlation

- $C(B_2)$ fails to capture the correlation
- $C(A_2)$ succeeds to detect the nonlinear correlation

Summary:
We have generalized the notion of covariance using group theory.

covariance → **theoretical properties** → **generalization**

$$\langle (x-\bar{x})(y-\bar{y}) \rangle$$

where

$$\langle \bullet \rangle = \int \bullet \, p(x,y)dxdy$$

2-variate cross cumulant

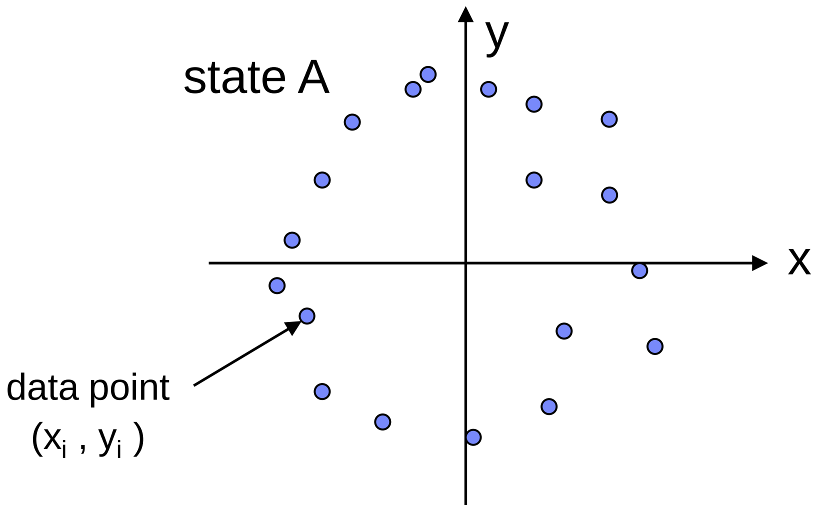
B_2 irreducible representation of C_{4v}

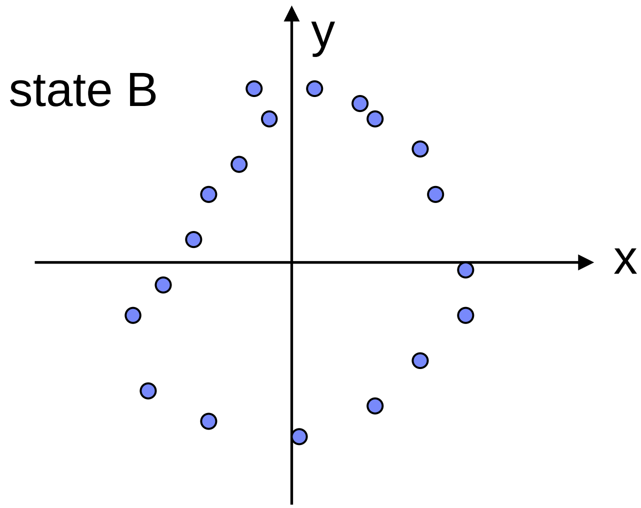
2-variate CC as A_1, A_2, B_2, and E representations

Symmetry decomposition:
new view to pattern recognition

Thank you !!

Background. How can you tell the difference of the two states quantitatively? The traditional covariance is not helpful.





- **Traditional covariance**

$$C_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

↑ # of data points

  - C_{xy} would be *useless* in this case.

- **The traditional covariance cannot capture nonlinearities.**

- **We wish to explicitly define useful metrics for nonlinear correlations.**
  - cf. kernel methods are black-boxes

Covariance as the lowest order CC: Summary of this section.
We focus on cross cumulants (CC) as a theoretical basis.

assumption

Describe *n*-variate systems using its PDF *p*

$$p(\mathbf{x}) = p(x_1, x_2, ..., x_n)$$

→ **characterization**

employ cumulants as features of *p*

→ **approximation**

- Take only 2-variate CC like $\langle {x_i}^{\mu} {x_j}^{\nu} \rangle_c$
- Take only lower order ones up to $\mu + \nu \leq 4$

Cumulant generating function

$$\Psi(\boldsymbol{s}) \equiv \ln \int d\boldsymbol{x}\, p(\boldsymbol{x}) \exp(\boldsymbol{s}^{\mathrm{T}} \boldsymbol{x}) = \ln \langle \exp(\boldsymbol{s}^{\mathrm{T}} \boldsymbol{x}) \rangle$$

Definition of cumulants

$$\Psi(\boldsymbol{s}) = \sum_j s_j \langle x_j \rangle_c + \dots + \frac{1}{k!} \sum_{j_1, \cdots, j_k} s_{j_1} \cdots s_{j_k} \langle x_{j_1} \cdots x_{j_k} \rangle_c + \dots$$

cumulants are expansion coeff. w.r.t. *s*

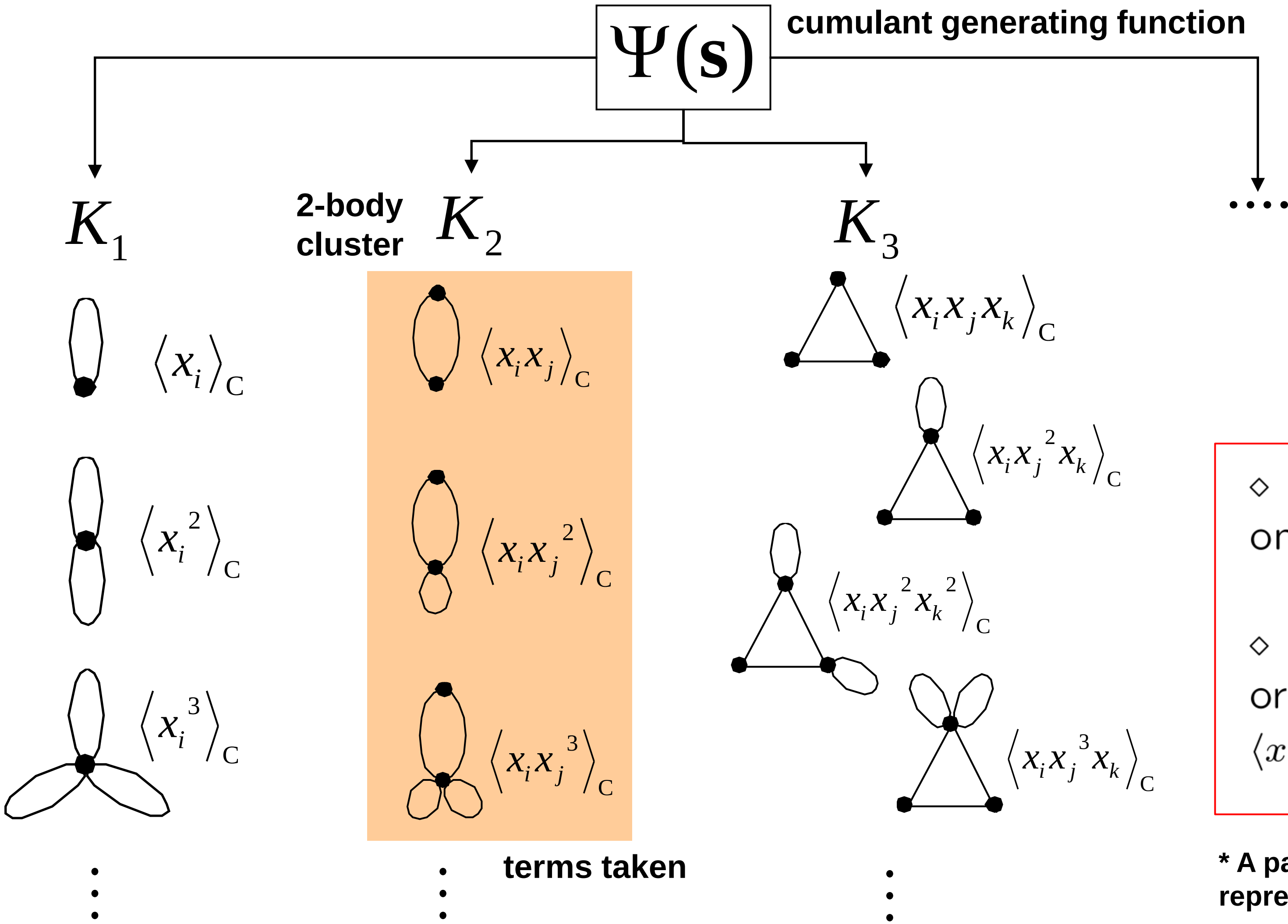
<u>Notation of "cumulant average"</u>

$$\langle {x_{i_1}}^{a_1} {x_{i_2}}^{a_2} \cdots {x_{i_k}}^{a_k} \rangle$$

↓

$(a_1, ..., a_k)$-th order cumulant w.r.t $(x_1, ..., x_k)$

Sparse Correlation Approximation of the cumulant generating function. What kind of terms are omitted?



SCA

- ◇ Approximate Ψ only with K_1 and K_2.
- ◇ Use only lower order terms in K_2: $\langle x^\mu y^\nu \rangle_C$, $\mu + \nu \leq 4$

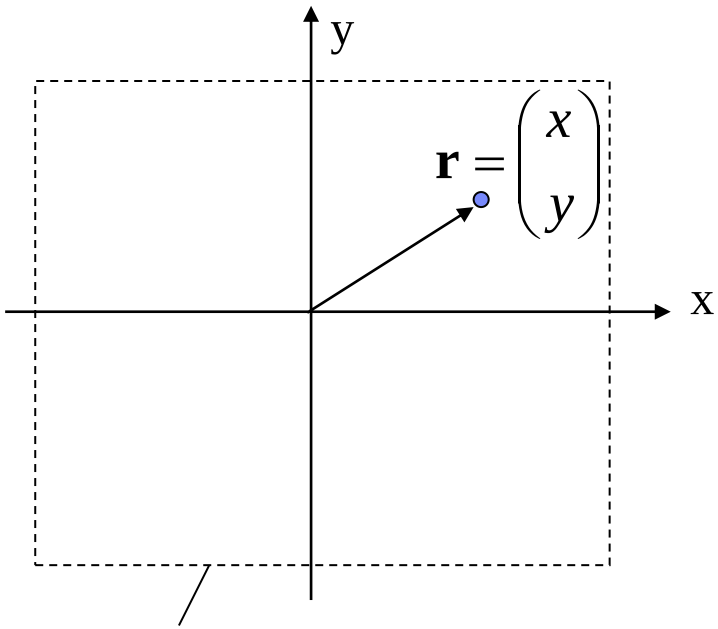
*** A pair of variable will be represented as *x* and *y* hereafter.**

Mathematical preliminaries. Position operator, Hilbert space, Dirac's bra-ket notation, and moments in the bra-ket notation

Def. of the position eigenstate

$$\mathbf{r} = \begin{pmatrix} a \\ b \end{pmatrix} \quad \text{then} \quad \hat{x}|\mathbf{r}\rangle = a|\mathbf{r}\rangle \quad \text{etc.}$$

position operator



D: The domain of *p*(***r***)

Hilbert space spanned by the position eigenstate

$$H = \{|\mathbf{r}\rangle \,|\, \mathbf{r} \in D\}$$

Def. of a state vector $|p\rangle$: $\quad |p\rangle = \int_D d\mathbf{r} |\mathbf{r}\rangle p(\mathbf{r})$

where $p(\mathbf{r}) = \langle \mathbf{r} | p \rangle$ is the marginal DF wrt (x,y)

Def. of a state vector $|x^\mu y^\nu\rangle$:

$$\langle \mathbf{r} | x^\mu y^\nu \rangle = x^\mu y^\nu$$

Bra-ket notation of moments

$$\langle p | x^\mu y^\nu \rangle = \int_D d\mathbf{r} x^\mu y^\nu p(\mathbf{r}) = \langle x^\mu y^\nu \rangle$$

Verifying the definition of the generalized covariances: a few examples.

(1) $|x^2y^2\rangle \in \mathcal{H}$

For $\forall g \in$ C$_{4\mathrm{v}}$, $\langle \boldsymbol{r}|g|x^2y^2\rangle = x^2y^2$.

representation matrices are all 1

→ A1 representation

$\mathrm{C}_{4\mathrm{v}}$	e	$C_4, {C_4}^3$	C_2	σ_x, σ_y	σ_ξ, σ_η
A_1	1	1	1	1	1
A_2	1	1	1	-1	-1
B_1	1	-1	1	1	-1
B_2	1	-1	1	-1	1
E	2	0	-2	0	0

Table 2. The character table of the C$_{4\mathrm{v}}$ group.

(2) $|xy^2\rangle$ and $|x^2y\rangle$

$\langle \boldsymbol{r}|C_4|xy^2\rangle = x^2y$ and $\langle \boldsymbol{r}|C_4|x^2y\rangle = -xy^2$

Thus, the representation matrix of C_4 is

$$D(C_4) = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},$$

so that we have $\mathrm{Tr}D(C_4) = 0$.

...

You can use the method of projection operators if want to construct IRRs systematically

(See a textbook of group theory)

Experiment : Calculating the generalized covariances for Lissajous' trajectories analytically.

- **Model of correlated variables →**
  - assume the uniformly distribution over *t*
  - mean is zero for both x and y
  - variance is 1 for both x and y

$$x(t) = \sqrt{2}\cos(\omega_1 t + \alpha)$$
$$y(t) = \sqrt{2}\sin(\omega_2 t + \beta)$$

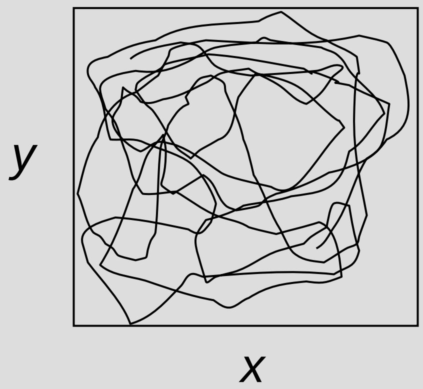
- **Generalized covariances can be explicitly calculated**

$$\begin{aligned}
\mathrm{C}(\mathrm{B}_2) &= \delta_{\omega_1,\omega_2}\sin\Omega_1^{\beta,\alpha} \\
\mathrm{C}(\mathrm{E}_1) &= -\frac{\delta_{\omega_1,2\omega_2}}{\sqrt{2}}\cos\Omega_2^{\alpha,\beta} + \frac{\delta_{2\omega_1,\omega_2}}{\sqrt{2}}\sin\Omega_2^{\beta,\alpha} \\
\mathrm{C}(\mathrm{E}_2) &= -\frac{\delta_{\omega_1,2\omega_2}}{\sqrt{2}}\cos\Omega_2^{\alpha,\beta} - \frac{\delta_{2\omega_1,\omega_2}}{\sqrt{2}}\sin\Omega_2^{\beta,\alpha} \\
\mathrm{C}(\mathrm{A}_1) &= -\frac{\delta_{\omega_1,\omega_2}}{2}\left[1 + 2\sin^2(\Omega_1^{\alpha,\beta})\right] \\
\mathrm{C}(\mathrm{A}_2) &= \frac{\delta_{\omega_1,3\omega_2}}{4}\sin\Omega_3^{\alpha,\beta} - \frac{\delta_{3\omega_1,\omega_2}}{4}\sin\Omega_3^{\beta,\alpha},
\end{aligned}$$

where we used the symbol $\Omega_c^{a,b} = a - bc$.

Cs are zero unless ω_1/ω_2 takes special values.

◇ If not, the xy space will be filled with the trajectory in the limit $t \to \infty$



Detailed summary.

- **We generalized the traditional notion of covariance based on the two theoretical properties**
  - 1. Standard covariance is the lowest order 2-variate cross cumulant
  - 2. Standard covariance is the B_2 irreducible representation of the C_{4v} group.
- **Our result suggests a new approach to pattern recognition where patterns are characterized by the irreducible representation of a finite group**
- **Practically, we found that**
  - $C(B_2)$ would be greatly enhanced for linear correlations.
  - $C(E_1)$ and $C(E_2)$ reflect some asymmetries in the distribution.
  - $C(A_1)$ clearly takes a large value when the distribution has a donut-like shape.
  - Finally, $C(A_2)$ would be enhanced by distributions with some Hakenkreuz-like correlations.
- **These features can be used in anomaly detection tasks where nonlinear correlations plays some important role**