Representations of finite groups and wireless communication

Vít Tuček

Huawei Sweden R&D, Stockholm

43rd WINTER SCHOOL GEOMETRY AND PHYSICS, 2023

Outline

Joint work with...

Collaborators = Huawei ∪ Cantabria

Huawei = {Olivier Verdier, Gunnar Peters}
Cantabria = {Diego Cuevas, Javier Álvarez-Vizoso, Carlos Beltrán, Ignacio Santamaría}

Huawei = Huawei Technologies Sweden AB
Cantabria = University of Cantabria

Section 1

Light introduction to wireless communication

Basic wireless communication model

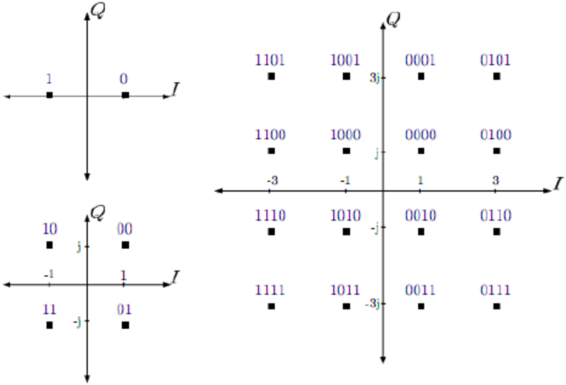
electromagnetic waves have amplitude and phase $\rightsquigarrow$ signals are modeled using complex numbers

M	number of transmit antennas
$X \in \mathbb{C}^{1\times M}$	transmitted symbol
N	number of receiving antennas
$Y \in \mathbb{C}^{1\times N}$	received symbol
$H \in \mathbb{C}^{M\times N}$	channel matrix (captures the propagation through environment)
$Z \in \mathbb{C}^{N}$	white Gaussian noise (i.e. iid $\mathbb{C}\mathcal{N}(0,\sigma^2)$)
ρ	signal to noise ratio (SNR) $\rho = \lVert X\rVert/\sigma$

$$Y = XH + Z \tag{1}$$

How does one actually send data?

Pick signals X only from a finite set $\mathcal{C}$ (ideally of size 2^B) and given received Y give a best guess as to which X could have produced it given our current knowledge of H.



The channel matrix H

- Depends on frequency and time of the transmission

$$H\colon I_t \times I_f \to H(t,f) \in \mathbb{C}^{M \times N}$$

- Numerous models (e.g. coming from the Maxwell equations) but the baseline is the so called Rayleigh fading where we assume $H(f,t)_{i,j} \sim \mathbb{C}\mathcal{N}(0,1) = \mathcal{N}(0,1_2)$.
- Can contain important correlations...

$$R_t = \mathbb{E}[HH^*] \in \mathbb{C}^{M \times M}, \quad R_r = \mathbb{E}[H^*H] \in \mathbb{C}^{N \times N}.$$

Estimation problems

At the beginning of the communication the receiver does not know the channel matrix H . . .

1. the communication protocol dictates that each communication begins with known pilot symbols $X_{p_1}, \ldots, X_{p_s}$
2. use pilot symbols to estimate H, R_t, R_r

But for high number of antennas this might be prohibitively expensive to do for each time and frequency!

Interpolation problem

Interpolate / extrapolate H, R_t in time and/or frequency domain.

Geometry in estimation

The covariance matrices have constraints (by definition)...

$$R_t \in Cov(M)$$
$$Cov(M) = \{A \in \mathbb{C}^{M \times M} \mid A^* = A \,\&\, \mathrm{spec}(A) \geq 0\}$$
$$\mathrm{GL}_M(\mathbb{C})/U(M) \subset Cov(M)$$

What is a "correct geometry" for the problem?
What about degenerate situations?

Precoding / beamforming

If the transmitter has access to the channel matrix H, we can improve the quality of the transmission:

$$X \rightsquigarrow W_H(X)$$

Zero forcing:

$$X \rightsquigarrow X(HH^*)^{-1}H$$

MMSE:

$$X \rightsquigarrow X(HH^* + 1/\rho I_M)^{-1}H$$

Truncated polynomial expansion:

$$X \rightsquigarrow X \sum_j w_j (HH^*)^j H$$

SVD-based precoding

$$H = VDU^*$$
$$D = \mathrm{diag}(d_1, \ldots, d_{\min\{M,N\}})$$
$$d_1 \geq d_2 \geq \cdots d_{\min\{M,N\}}$$

Coordinates of X wrt basis of left singular vectors, which correspond to small singular values, are drowned by the noise.
If the transmitter knows the k largest singular vectors $(v_1, \ldots, v_k)$, it can use them for precoding and get better power efficiency / effective SNR.

$$X \rightsquigarrow X[v_1|\cdots|v_k]$$

Precoding geometry – $k = 1$

Left singular vectors are the eigenvectors of HH^* and hence they are defined up to nonzero complex multiple.
In other words:

$$\mathrm{Sing}_1 \simeq \mathbb{C}P^{M-1} \simeq U(M)/U(1) \times U(M-1)$$

Precoding geometry – $k \in \{1, \ldots, M\}$

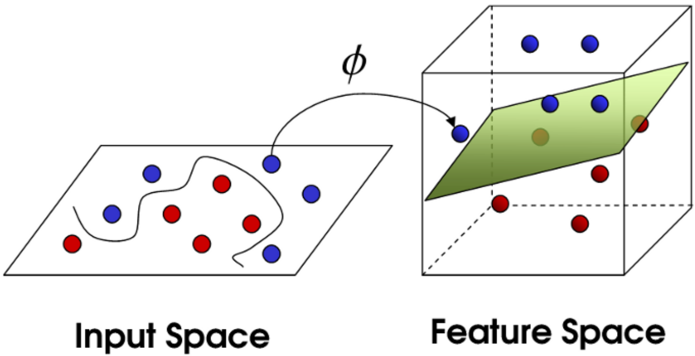
Generically,[2] the space of k singular vectors corresponding to k largest singular values is

$$\mathrm{Sing}_k \simeq U(M)/U(1) \times \cdots \times U(1) \times U(M-k)$$

[2] In case of multiple singular values we have

$$U(M)/U(k_1) \times \cdots \times U(k_s) \times U(M - \sum_{i=1}^{s} k_i)$$

Kernel based approach



Kernel based approach

How to linearize complicated space?

With reproducing kernel Hilbert space!

RKHS:

X space we are interested in

$\mathcal{H}$ Hilbert space

$\Phi\colon X \to \mathcal{H}$ feature map

such that there exists $k\colon X \times X \to \mathbb{C}$ with the property:

$$\forall x, y \in X \quad k(x, y) = \langle \Phi(x) | \Phi(y) \rangle_{\mathcal{H}}$$

Any finite computation involving just the scalar product can be done by evaluating the kernel function.

Representation theory to the rescue?

Fix a closed subgroup $K \leq G$ and a unitary G-representation $\mathcal{H}$.

For any $v_0 \in \mathcal{H}^K$ the closed G-invariant subspace $\mathcal{H}_0$ generated by v_0 is RKHS which is realized on $\mathcal{C}(G/K)$.

Remark: For applications, we care only about effective algorithm for evaluating the kernel function with "good enough" numerical precision.

Possible future project, not yet approved. :-/

Section 2

Grassmannian communication

Block fading

Assume that the channel matrix does not change for T transmissions:

$$
\begin{aligned}
Y_1 &= X_1 H + Z_1 \\
&\vdots \\
Y_T &= X_T H + Z_T
\end{aligned}
$$

$$Y = XH + Z$$

$X \in \mathbb{C}^{T \times M}$ transmitted symbol
$Y \in \mathbb{C}^{T \times N}$ received symbol
$H \in \mathbb{C}^{M \times N}$ channel matrix (captures the propagation through environment)
$Z \in \mathbb{C}^{T \times N}$ white Gaussian noise (i.e. iid $\mathbb{C}\mathcal{N}(0, \sigma^2)$)

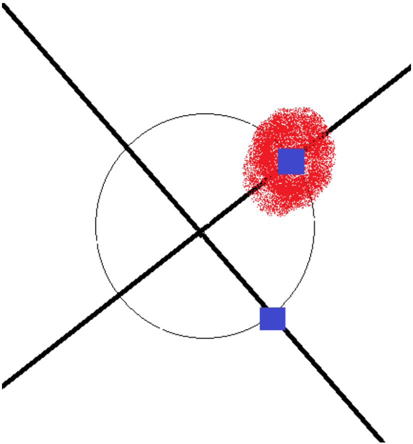
Grassmannian communication

If $M = N$ and $Z = 0$ then we have

$$Y = XH$$

and

$$\mathrm{colspan}(X) = \mathrm{colspan}(Y).$$

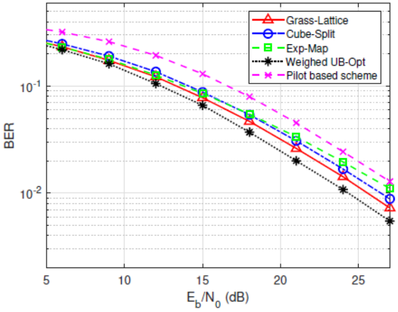


Classical vs Grassmannian signaling

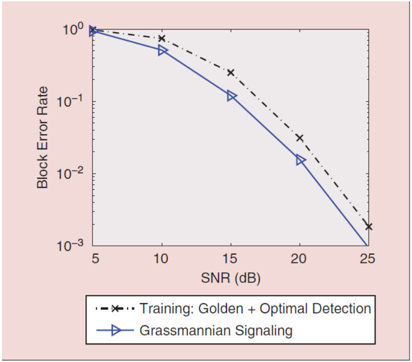
Degrees of freedom:

classicaly $X \in \mathbb{C}^{T \times M}$ ... MT

Grassmannian $X \in Gr(M, T)$ ... $TM - M^2 = M(T - M)$



64 points on $Gr(1, 2)$



4096 points on Gr(2, 4)

Details of Grassmannian signaling

We assume that H is iid $\mathbb{C}\mathcal{N}(0,1)$ (Rayleigh block fading).
We start with the conditional probability of receiving $Y \in \mathbb{C}^N$ assuming $X \in \mathbb{C}^M$ was sent.

$$\mathbf{Y} = X\mathbf{H} + \mathbf{Z}$$

$$P(Y|X) = \frac{\exp(-\operatorname{tr}(Y^*(1_T + XX^*)^{-1}Y))}{\pi^{TN}\det(1_T + XX^*)}$$

Observation:

$$\forall h \in U(M): \; P(Y|Xh) = P(Y|X)$$
$$\forall g \in U(T): \; P(gY|gX) = P(Y|X)$$

Capacity

Theorem (Marzetta-Hochwald-Zheng-Tse-Durisi-Riegler)

Assume $T \geq M + N$. Given a constraint on power of the signal (e.g. $\|X\|_F = 1$) the distribution on X that maximizes the Shannon information $I(Y;X)$ is the uniform distribution on the Grassmannian.

$$I(Y;X) = \mathbb{E}\ \log \frac{p(Y|X)}{p(Y)}$$

$$C = \sup_{p_X} I(Y;X)$$

Capacity

Proof.

1. $\forall g \in U(T) \forall h \in U(M): \; I(Y|X) = I(Y|g^{-1}Xh)$
2. Let p_0 be a fixed probability distribution of X and define

$$p_1(X) = \frac{1}{|U(T)||U(M)|} \int_{g \in U(T)} \int_{h \in U(M)} p_0(g^{-1}Xh).$$

Since $I(Y|X)$ is concave wrt p_X we have by the Jensen's inequality

$$I(Y|X_{p_1}) \geq I(Y|X_{p_0}).$$

□

Capacity

Proof.

3. Capacity achieving distribution $X = gD$ where g is uniformly distributed on $U(T)$ and independent of D which is $T \times M$ nonnegative diagonal matrix whose pdf is invariant with respect to permutations.
4. For $T \geq M + N$ we can drop the diagonal factor, for $T < M + N$ the capacity achieving distribution is nontrivial.

□

Detection

In practical situation we consider finite set of tall unitary matrices $X^*X = 1_M$.

$$\mathcal{C} = \{X_1, \dots, X_k\}$$

Given a received signal Y, how do we guess which X_i was sent?

Definition (Maximum Likelihood Detector)

$$ML(Y) = \arg\max_{X \in \mathcal{C}} p(Y|X)$$

Towards codebook criteria

$$P(Y|X) = \frac{\exp(-\operatorname{tr}(Y^*(1_T + XX^*)^{-1}Y))}{\pi^{TN}\det(1_T + XX^*)}$$

Since we assume $X^*X = 1_M$ we can simplify

$$ML(Y) = \arg\max_{X\in\mathcal{C}} \operatorname{tr}(YY^*XX^*).$$

Moreover, we can interpret XX^* as the orthogonal projection to the subspace of $\mathbb{C}^T$ spanned by the columns of X.

Grassmannians

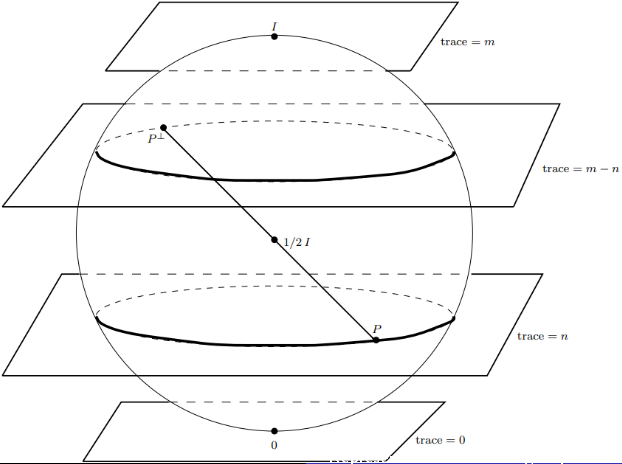
$$
\begin{aligned}
Gr(M,T) &= \{V \leq \mathbb{C}^T \mid \dim V = M\} \\
&\simeq U(T)/U(M) \times U(T-M) \\
&\simeq \{X \in M(T,M,\mathbb{C}) \mid X^*X = 1_M\}/U(M) \\
&\simeq \{P \in M(T,T,\mathbb{C}) \mid P^* = P \,\&\, P^2 = P \,\&\, \mathrm{rank}\, P = M\} \\
&= \{P \in \mathrm{Sym}(T) \mid P^2 = P \,\&\, \mathrm{tr}\, P = M\}
\end{aligned}
$$

Frobenius inner product on $\mathrm{Sym}(T)$:

$$
\langle A \mid B \rangle_F = \mathrm{tr}(A^* B)
$$

Embedings of Grassmannians

J.H. Conway, R.H. Hardin, N.J.A. Sloane: *Packing Lines, Planes, etc.: Packings in Grassmannian Space*



Grassmannians

Definition (Chordal distance)

$$d_{Ch}(A, B) = \|A - B\|_F$$

On $Gr(M, T)$ this restricts to

$$d_{Ch}(A, B) = \sqrt{2}\sqrt{M - \operatorname{tr}(AA^*BB^*)}$$

and so

$$ML(Y) = \arg\min_{X \in \mathcal{C}} d_{Ch}(YY^*, XX^*)$$

Towards codebook criteria

Problem

What is the optimal constellation $\mathcal{C} = \{X_1, \ldots, X_k\}$ of a given size?

Since our ML detector is picking up the closest constellation point wrt the chordal distance a good choice might be

Chordal criterion

$$\mathcal{C}_{ch} = \arg\max_{\mathcal{C}} \min_{X_i \neq X_j \in \mathcal{C}} d_{ch}(X_i, X_j)$$

but can we justify that?

Towards codebook criteria – pairwise error

Pairwise error of mistaking X_i for X_j is

$$P_e(X_i, X_j) = \sum_{j=1}^{M} \mathrm{Res}_{w=\imath a_j} \left(\frac{-1}{w + \imath/2} \prod_{m=1}^{M} \left(\frac{1+\alpha}{\alpha^2(1-d_m^2)(w^2+a_m^2)} \right) \right)$$

where $\alpha = \rho T/M$, $a_j^2 = 1/4 + (1+\alpha)/(\alpha^2(1-d_j^2))$ and $1 \geq d_1 \geq d_2 \cdots \geq d_m$ are the singular values[3] of $X_j^* X_i$. The product omits the terms where $d_m = 1$.

Theorem (Cuevas-Santamaria-T)

$$P_e(X_i, X_j) = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \prod_{m=1}^{M} \left(1 + \frac{\alpha^2(1-d_m^2)}{4(1+\alpha)\cos^2\theta} \right)^{-N} \mathrm{d}\theta$$

[3] the cosines of the principal angles between the subspaces $[X_i]$ and $[X_j]$

Towards codebook criteria – pairwise error

For $\rho \to 0$ we have

$$P_e(X_i, X_j) = 1/2 - \frac{T\sqrt{N} d_{ch}(X_i, X_j)}{4M} + o(\rho)$$

Chordal criterion

$$\mathcal{C}_{ch} = \arg\max_{\mathcal{C}} \min_{X_i \neq X_j \in \mathcal{C}} d_{ch}(X_i, X_j)$$

Towards codebook criteria – pairwise error

For $\rho \to \infty$ we have

$$\lim_{\rho\to\infty} \rho^{MN} P_e(X_i, X_j) = \frac{1}{2}\left(\frac{4M}{T}\right)^N M \frac{(2NM-1))!!}{(2NM)!!} \prod_{m=1}^{M} (1-d_m^2)^{-N}$$

Coherence criterion

$$\mathcal{C}_{coh} = \arg\max_{\mathcal{C}} \min_{X_i \neq X_j \in \mathcal{C}} \det(1_M - X_i^* X_j X_j^* X_i)^N$$

Union bound criterion

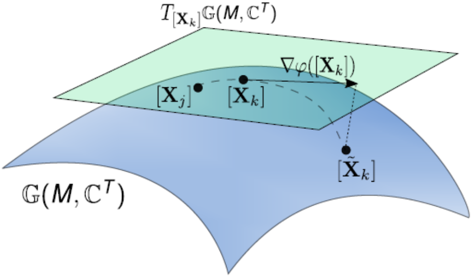
$$\mathcal{C}_{UB} = \arg\min_{\mathcal{C}} \sum_{i<j} \det(1_M - X_i^* X_j X_j^* X_i)^{-N}$$

Section 3

Constellations of subspaces

Constellation design - numerical optimization

1. Start with random constellation of the given number K of points.
2. At each iteration, for each point X_k find the L "closest points" and move X_k away from its neighbors.



Works nice but we want $|\mathcal{C}| = 2^B$...

GrassLattice

We can efficiently construct & detect on rectangular grids.

Problem

Can we map such a grid invertibly into the Grassmannian so that it is near optimal wrt our cost functions?

GrassLattice ($M = 1$)

1. Take the unit hypercube in $\mathbb{R}^{2(T-1)}$ and map[1] it through

$$(a_1, \dots, a_n, b_1, \dots, b_{T-1}) \mapsto (z_i = F^{-1}(a_i) + \imath F^{-1}(b_i))_{i=1}^{T-1} \in \mathbb{C}^{T-1}.$$

2. Map the $\mathbb{C}\mathcal{N}(0, 1_{T-1})$-distributed vector $z \in \mathbb{C}^{T-1}$ to the unit disc by

$$z \mapsto w = zf(\|z\|)$$

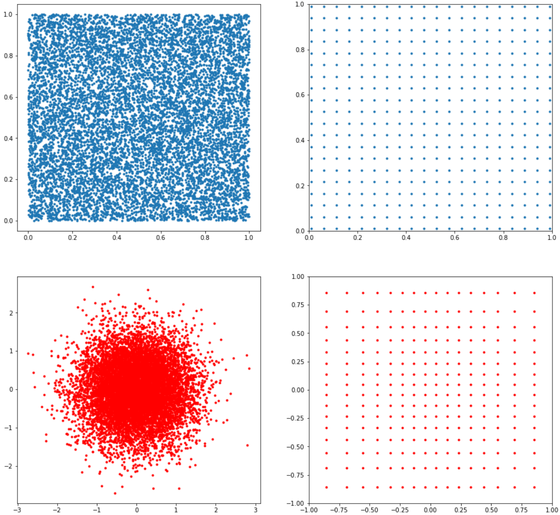
where $f(r) = \frac{1}{r}\left(1 - \exp(-t^2)\sum_{k=0}^{T-2} \frac{r^2 k}{k!}\right)^{1/2(T-1)}$
(it makes w is uniformly distributed in the unit disc $B(0,1) \in \mathbb{C}^n$)

3. Map w to the Grassmannian by

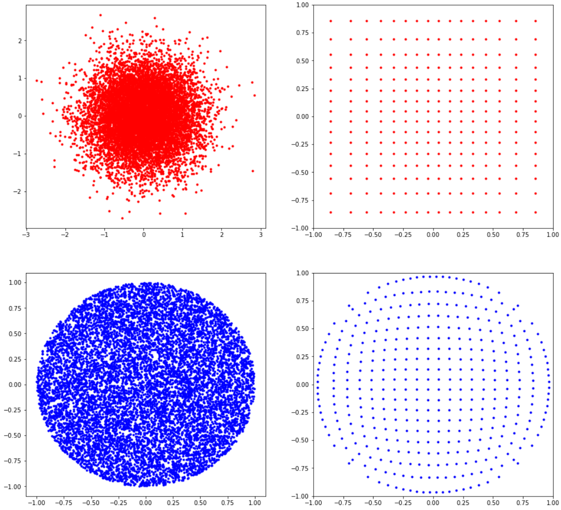
$$w \mapsto (\sqrt{1 - |w|^2}, w)$$

[1] F is the distribution function of $\mathcal{N}(0, 1/2)$

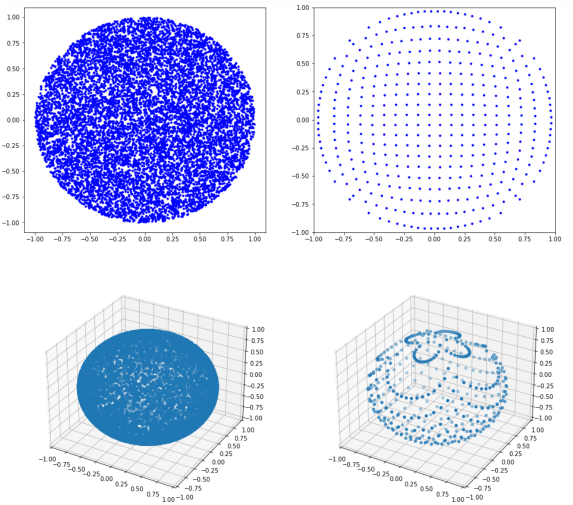
GrassLattice ($M = 1$) in pictures



GrassLattice ($M = 1$) in pictures

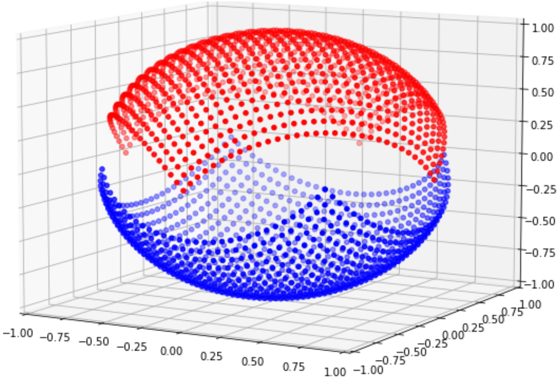


GrassLattice ($M = 1$) in pictures



GrassLattice ($M = 1$) – the other chart?

We can alternatively use $w \mapsto (w, \sqrt{1 - |w|^2})$ and get twice as many points if we shrink the lattice and rotate one chart



What rotations should one choose for $T > 2$?

GrassLattice for $M > 1$?

The last map $w \mapsto (\sqrt{1 - |w|^2}, w)$ is actually not just measure preserving but even symplectomorphism.

For general M, the map

$$W \mapsto \begin{pmatrix} \sqrt{1_M - W^* W} \\ W \end{pmatrix}$$

is also symplectomorphism map into $Gr(M, T)$ from the set of matrices where the square root is well defined:

$$\{W \in \mathbb{C}^{(T-M)\times M} \mid \|W\|_{op} < 1\} \text{ (Cartan domain of type I)}$$

GrassLattice for $M > 1$?

Problem

Let D be a Cartan symmetric domain of type I.

Can we explicitely map a unit hypercube into D in a measure-preserving way?

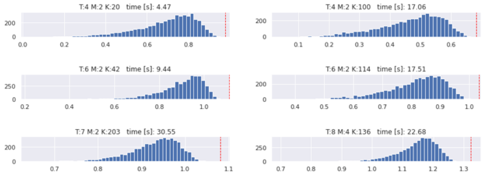
Given $B \in \mathbb{N}$, can we efficiently construct 2^B points in D so that the resulting subspace constellation is close to optimal?

Finite group constellations

Given a finite subgroup $G \leq U(T)$ and a basepoint $[B] \in Gr(M, T)$ we can consider its orbit as a constellation

$$\mathcal{C}_{G,B} = \{[gB] \mid g \in G\}$$

- basepoint matters
- generically $|C_{G,B}| = |G|$ but smaller orbits can be also useful



Example [Pitaval, Tirkkonen]

The following basepoint is optimal for two dimensional representation of the dihedral group D_8 giving rise to a constellation of 8 points on $\mathbb{C}P^1$.

$$\begin{pmatrix} \cos\frac{1}{4}\arccos(\frac{3}{7}-\frac{6\sqrt{2}}{7}) \\ \left(\frac{1}{2^{1/4}}+\imath\sqrt{1-\frac{1}{\sqrt{2}}}\right)\sin\frac{1}{4}\arccos(\frac{3}{7}-\frac{6\sqrt{2}}{7}) \end{pmatrix}$$

Finite group constellations – finding good basepoint

1. Instead of optimizing over $\prod_{k=1}^{K} Gr(M, T)$ we optimize just over $Gr(M, T)$.
2. Our cost functions are $U(T)$-invariant which reduces the evaluation complexity from K^2 to K:

$$\{d_{ch}([g_i B], [g_j B]) \,|\, (g_i, g_j) \in G\} = \{d_{ch}([gB], [B]) \,|\, g \in G\}$$

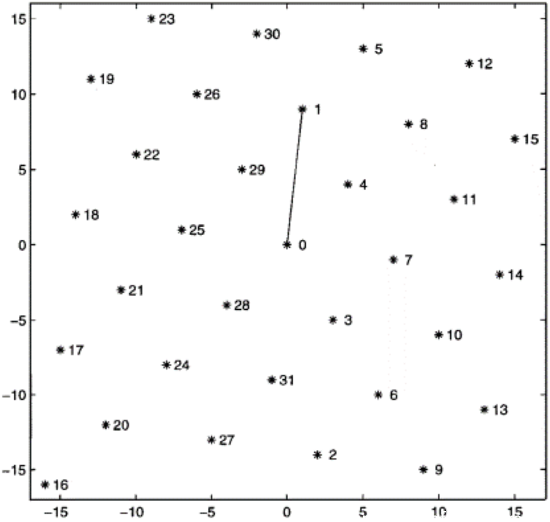
3. Further simplifications:

Criteria for group-based constellations

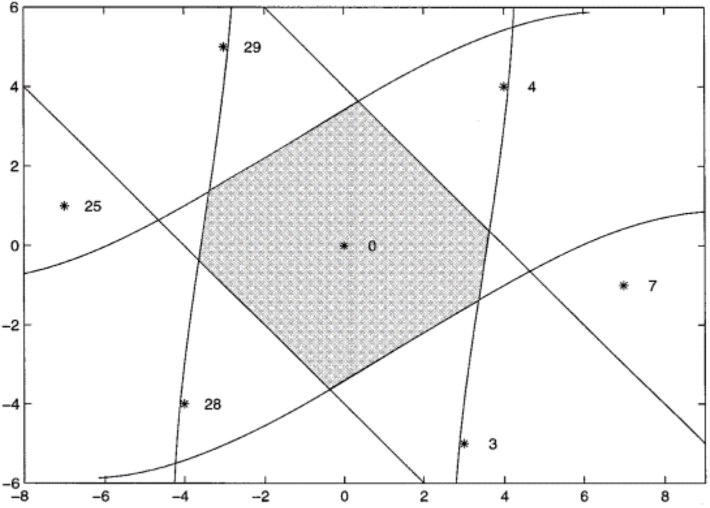
$$\mathcal{C}_{ch} \longleftarrow \underset{[B]\in Gr(M,T)}{\arg\min} \max_{g\in G} \mathrm{Tr}[(B^* g B)(B^* g^{-1} B)]$$

$$\mathcal{C}_{UB} \longleftarrow \underset{[B]\in Gr(M,T)}{\arg\min} \sum_{g\in G} \det[1_M - (B^* g B)(B^* g^{-1} B)]^{-N}$$

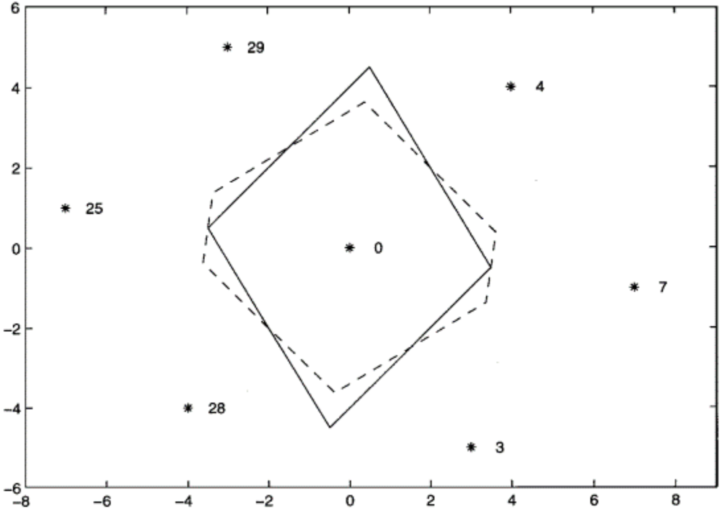
Detection for abelian subgroups



Detection for abelian subgroups



Detection for abelian subgroups



Finite group constellations – advantages

1. encoding and storage:
   $G = \{g_1^{i_1} \cdots g_k^{i_k} \mid i_j = 1, \ldots, N_j\}$ and $k \sim \log |G|$
2. for each group one gets constellation for any Grassmannian (add transmit antennas = store one more basepoint)
3. some provably optimal constellations are of this type (see e.g. Conway et al)

Finite group constellations – performance

Finite group constellations – obstacles

Problem

Which subgroups should we choose?

For our applications we could in principle just numerically explore finite subgroups of $U(T)$ for $T \leq 10$, but classification of finite subgroups of $U(T)$ is known only for $T \leq 4$

Finite group constellations – obstacles

Theorem (Jordan)

There exists a real function f such that every finite subgroup of $GL_d(\mathbb{C})$ has a normal Abelian subgroup of index bounded by $f(d)$.

$$f(d) = (d+1)! \text{ for } d \geq 71$$

Finite group constellations – group approximability

Let $\epsilon > 0$. Consider a metric group G with a left-invariant distance function. We say that G is *ϵ-approximable* if there exists a finite *subset* $H \subset G$ and with its own group law $\circ_H$ such that

1. For each $g \in G$ there exists a point in H of distance at most ϵ.
2. For each $a, b \in H$ we have $d(a \circ_G b, a \circ_H b) \leq \epsilon$.

Group G is approximable if it is ϵ-approximable for any $\epsilon > 0$.

Finite group constellations – obstacles

Theorem (Turing)

1. *If a metric group is approximable and has a faithfull representation in $GL(\mathbb{C}, d)$, then it is approximable by groups which also have faithful degree d linear representations.*
2. *If a Lie group is approximable, then it is compact and abelian.*

Thank you!