Experimenting with the Gram-Schmidt Walk

Gaëtan Bossy
Supervision by Pr. Adam Marcus
Chair of Combinatorial Analysis

Banaszczyk's Theorem

A fondamental theorem in discrepancy theory is the following result:

Theorem 1 (Banaszczyk, 1998) *For all convex body $K \subseteq \mathbb{R}^d$, with Gaussian measure $\gamma_m(K) \geq 1/2$, and given $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in \mathbb{R}^d$, $\|\boldsymbol{v}_i\|_2 \leq 1$ for all i, then there exists $\boldsymbol{z} \in \{-1,1\}^n$ such that $\sum_{i=1}^n \boldsymbol{z}(i)\boldsymbol{v}_i \in 5K$.*

While this results gives the best known bounds for several well-known discrepancy problems, its proof is non-constructive. For years, mathematicians tried to come up with a constructive algorithm to generate the colorings whose existence is proved by the theorem. In 2016, Dadush, Garg, Lovett and Nikolov proved that actually, all that is needed is an algorithm to generate a coloring $\mathbf{z}$ such that the corresponding **vector of imbalances**, $\mathbf{Bz}$, where $\mathbf{B} = (\mathbf{v}_1, \dots, \mathbf{v}_n) \in \mathbb{R}^{d\times n}$, would be σ-subgaussian with $\sigma > 0$ a constant.

The Gaussian measure of a body is defined as

$$\gamma_m(S) = \mathbb{P}[\mathbf{g} \in S]$$

where $\mathbf{g}$ is a standard Gaussian random vector in $\mathbb{R}^m$, i.e. $\mathbf{g} \sim \mathcal{N}(0, I_m)$.

A random vector $\mathbf{Y} \in \mathbb{R}^m$ is said to be subgaussian with parameter σ (or σ-subgaussian) if for all $\boldsymbol{\theta} \in \mathbb{R}^m$:

$$\mathbb{E}\left[e^{\langle \mathbf{Y}, \boldsymbol{\theta}\rangle}\right] \leq e^{\sigma^2\|\boldsymbol{\theta}\|_2^2/2}.$$

The Gram-Schmidt Walk Algorithm

Algorithm 1: The Gram-Schmidt walk (**GSW**) by Dadush, Bansal, Garg and Lovett, 2018

Input : $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^d$, an initial coloring $\mathbf{z}_0 \in [-1,1]^n$

Output: a coloring $\mathbf{z}_n \in \{-1,1\}^n$

1 $t = 1, A_1 = \{i \in [n] : |\mathbf{z}_0(i)| < 1\}$ and $p(1) = \max\{i \in A_1\}$

2 **while** $A_t \neq \emptyset$ **do**

3 $\quad \mathbf{u}_t = \arg\min_{\mathbf{u}\in U} \|\mathbf{Bu}\|$ where $U = \{u \in \mathbb{R}^n : u(p(t)) = 1 \text{ and } u(i) = 0 \forall i \notin A_t\}$

4 $\quad \Delta = \{\delta : \mathbf{z}_{t-1} + \delta\mathbf{u}_t \in [-1,1]^n\}$, let $\begin{cases} \delta_t^+ = \max\Delta \\ \delta_t^- = \min\Delta \end{cases}$ then $\delta_t = \begin{cases} \delta_t^+ \text{ w.p. } \frac{-\delta_t^-}{(\delta_t^+ - \delta_t^-)} \\ \delta_t^- \text{ w.p. } \frac{\delta_t^+}{(\delta_t^+ - \delta_t^-)} \end{cases}$

5 $\quad \mathbf{z}_t = \mathbf{z}_{t-1} + \delta_t\mathbf{u}_t$, $t \leftarrow t+1$, $A_t = \{i \in [n] : |\mathbf{z}_{t-1}(i)| < 1\}$, $p(t) = \max\{i \in A_t\}$

6 **end**

7 Output $\mathbf{z}_T \in \{-1,1\}^n$.

This is equivalent to choosing the pivot randomly out of A_t when the previous pivot has been colored.

If $V_t = \text{span}\{\mathbf{v}_i : i \in A_t \setminus \{p(t)\}\}$, $\mathbf{u}_t$ can be chosen as a solution of $\Pi_{V_t}(\mathbf{v}_{p(t)}) + \sum_{i\in A_t\setminus\{p(t)\}} \mathbf{u}_t(i)\mathbf{v}_i = \mathbf{0}$, $\mathbf{u}_t \in U$ where Π_{V_t} is the projector on V_t.

$(\mathbf{z}_t)_{t\in\{1,\dots,n\}}$ is actually a martingale as $\mathbb{E}[\delta_t|\delta_t^+, \delta_t^-] = 0$.

This algorithm produces an assignment $\mathbf{z}$ such that the corresponding vector of imbalances, $\mathbf{Bz}$, has a small norm. Another way to describe it is that it divides the vectors in two groups such that the sum of the vectors in each group are pretty close. One small modification one can do is to always choose the δ with the smallest absolute value in line 4. We call this modification the **Deterministic Gram-Schmidt Walk** (**DGSW**).

Theorem 2 (Harshaw, Spielman, Zhang, Sävje, 2019) *For $\boldsymbol{z}$ sampled via the Gram-Schmidt walk with input $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n \in \mathbb{R}^d$ with euclidean norm at most 1, we have that $\boldsymbol{Bz}$ is subgaussian with parameter $\sigma^2 = 1$.*

Changing the Pivot Rule

In the pseudocode given above, the pivot $p(t)$ is chosen through the order of the input vector. It could also be chosen randomly, as that would be equivalent to shuffling the input vectors which do not have a specific order. One area of potential improvement for this algorithm is to add rules for the choice of the pivot. For example, it could depend on the norm of the vectors in A_t, on how much they would move the vector of imbalances in expectation, or on the fractional coloring. One promising variant that seems to improve the algorithm regarding the minimization of the vector of imbalances is to choose the pivot as $p(t) = \arg\max_{i\in A_t} |x(i)|$. This rule, called **maximum absolute coloring** (**MAC**) changes the behavior of the algorithm as can be seen in figures 3 and 1. The assignments produced using this rule result in notably shorter vector of imbalances, especially when coupled with the DGSW. In my thesis, we explore several other potential modifications of the algorithm.

Plot of Average Norm of Vectors of Imbalances

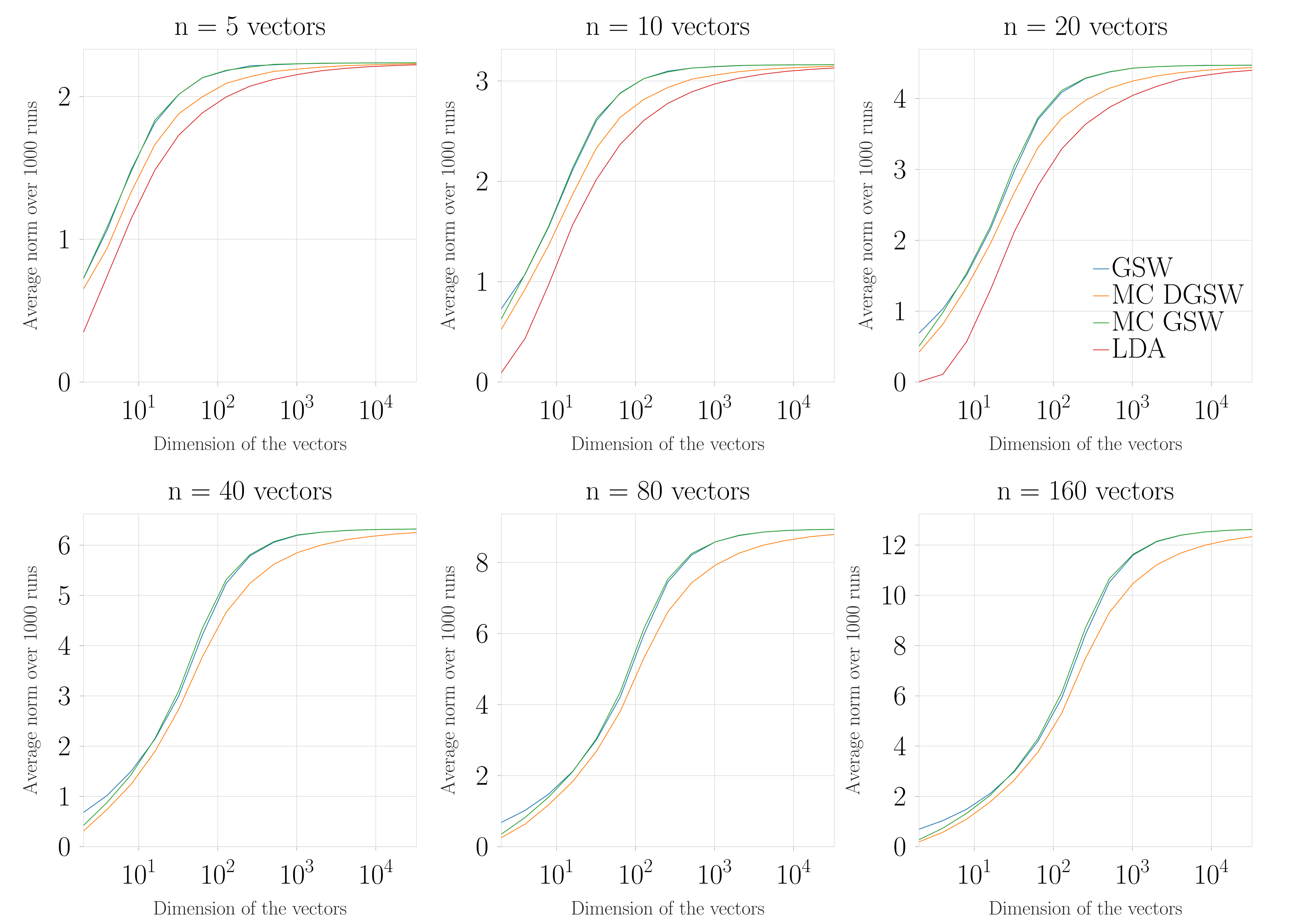

Fig. 1: Plot of average norm of vectors of imbalances for the GSW, the GSW with MAC and the DGSW with MAC, each obtained from an input of n vectors sampled uniformly from the ball of radius 1 in $\mathbb{R}^d$ with random pivot rule and maximum absolute coloring pivot rule. For small n's, there is also a plot of the average norm of shortest vectors of imbalances obtained by brute force, named as **Lowest Norm Assignment** (**LDA**). Results are averaged for 1000 runs except for the LDA with $n = 20$ where they're averaged over 20 runs as the running time is too long to do more of them.

A Complete Example for $n = 3, d = 2$

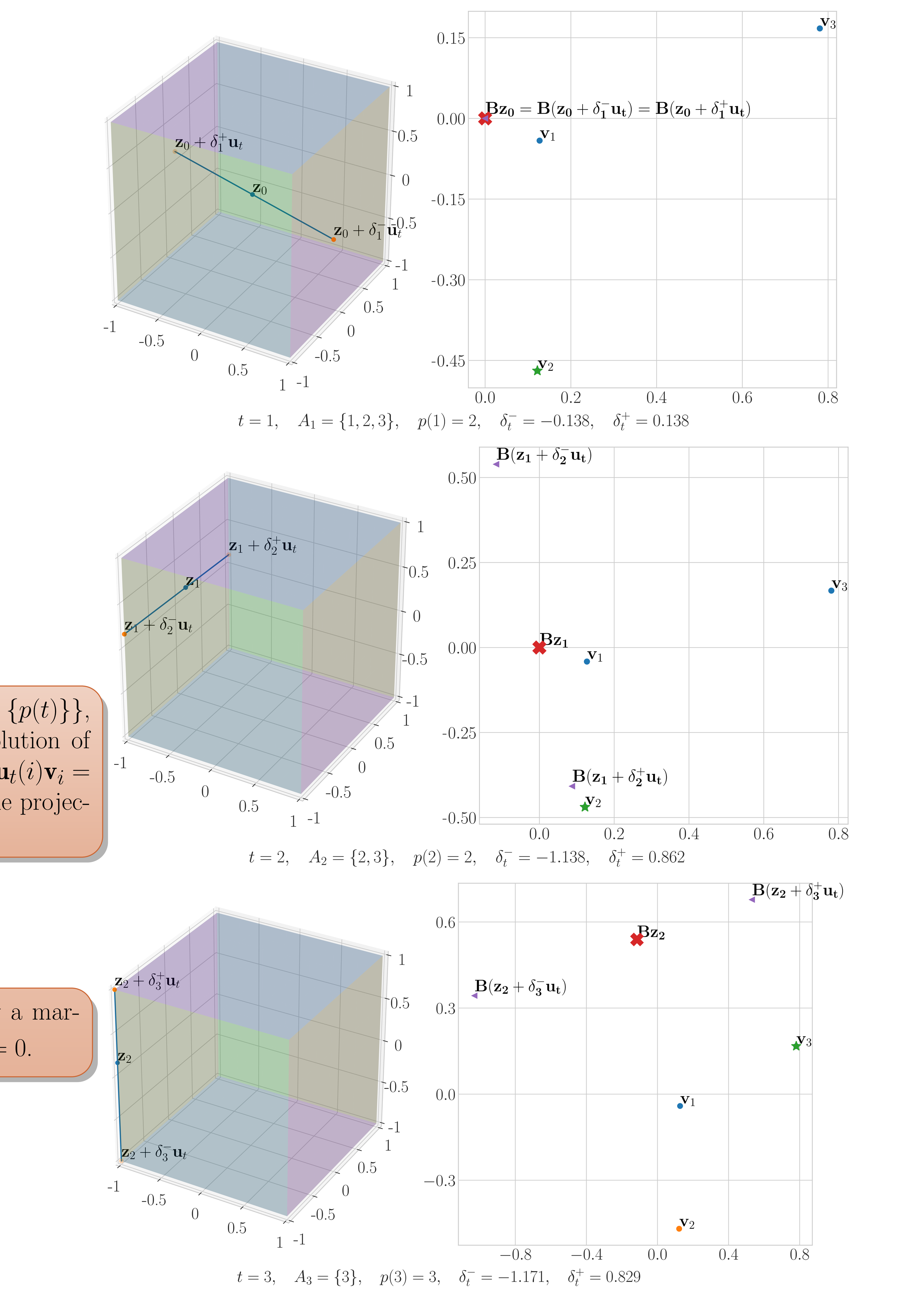

Fig. 2: Example of a GSW run with 3 vectors in $\mathbb{R}^2$. The left part shows the cube where the coloring is living, and the right part shows the vector of imbalances and the input vectors.

Plot of Vectors of Imbalances

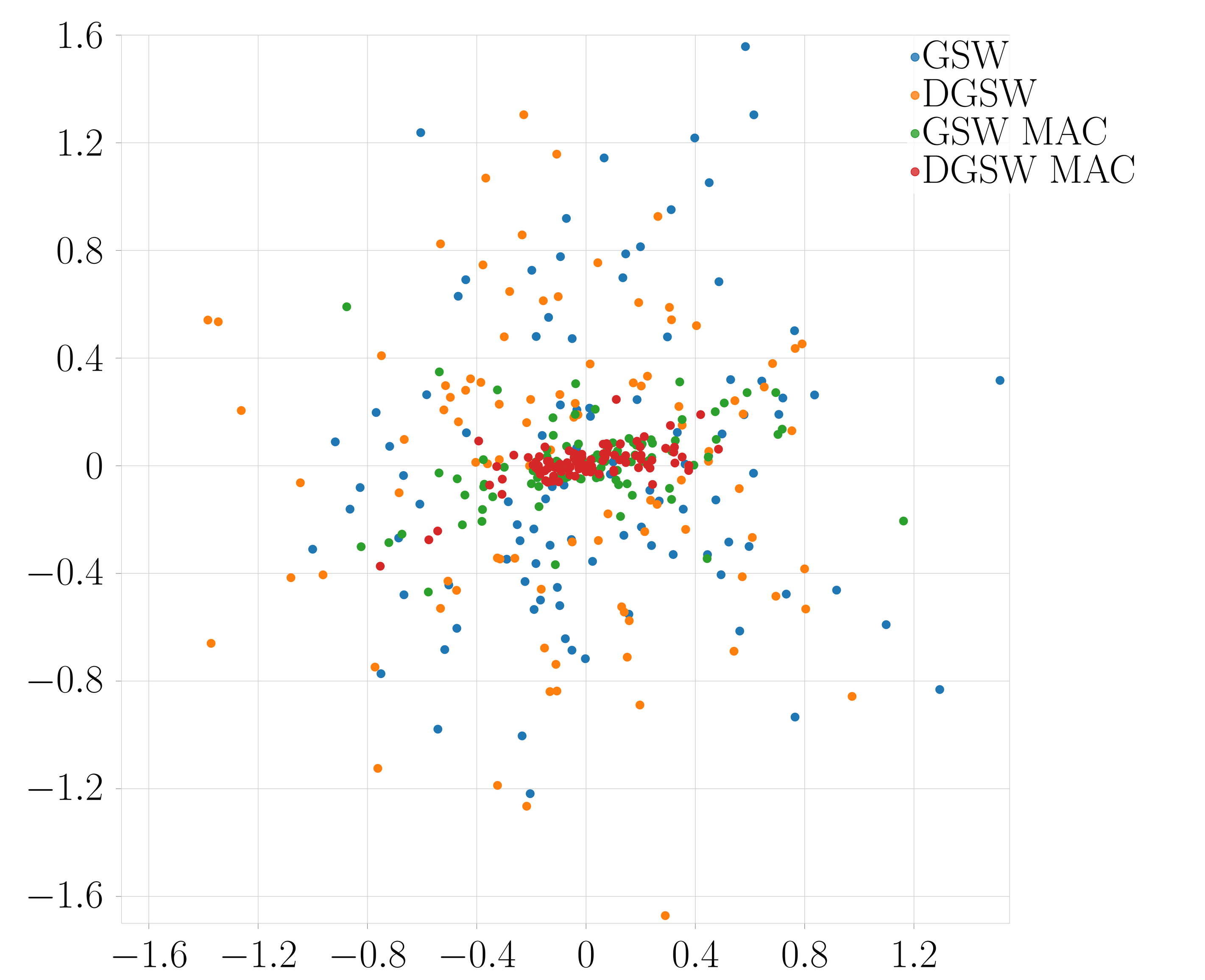

Fig. 3: Plot of vectors of imbalances for the GSW and DGSW, each obtained from an input of 100 vectors sampled uniformly from the ball of radius 1 in $\mathbb{R}^2$ with random and maximum absolute coloring pivot rules.

Going Further

There are various areas where this algorithm could be improved and better understood:

- The choice of the direction could be changed slightly to adapt it to some problematic inputs that exist for $n > d$.
- The pivot choice could be done in many ways to improve some aspects.
- The input vectors could be modified to add additional known information or constraints about the assignment.
- The modification discussed in my thesis, that minimizes balance in a certain way, could certainly be better understood.
- The algorithm could be generalized to separate inputs into more than 2 groups in a single run.

Additionally, it would also be interesting to explore further the use of the algorithm to solve problems regarding hidden substructures in large sets, such as the planted clique, or the potential usages of a vector balancing algorithm in machine learning.