Discrepancy: Lower Bound for the Greedy Algorithm

Given a set system $(X, \mathcal{R})$, a two-coloring of X, say denoted by χ, is an assigment of a '-1' or a '$+1$' color to each vertex of X. That is,

$$\chi\colon X \to \{-1, +1\} .$$

Then the *discrepancy* of $\mathcal{R}$ with respect to χ, is defined as:

$$\mathrm{disc}_\chi(\mathcal{R}) = \max_{S \in \mathcal{R}} \left| \sum_{v \in S} \chi(v) \right| .$$

Our goal is to compute a two-coloring of X with low discrepancy. Consider the following 'greedy' algorithm to compute such a coloring:

We will color the elements of X sequentially—in the order $v_1, \ldots, v_n$—with a '$+1$' or a '-1' color. At the start, each element is uncolored, and considered to have a color of 0.

Then for $i = 1, \ldots, n$:

1. Among all the sets of $\mathcal{R}$ containing v_i, pick the one with maximum discrepancy (pick an arbitrary one if several choices), and denote this set by S^i. That is,

$$S^i = \arg \max_{\substack{S \in \mathcal{R}: \\ v_i \in S}} \mathrm{disc}(S).$$

2. Assign v_i a color that *decreases* the discrepancy of S^i.

The result of this section is the following.

Theorem 0.59. *Given integers m and n with $m \geq n \geq 1$, there exists a set system $(X, \mathcal{R})$ with $|X| = n$ and $|\mathcal{R}| = m$, such that the coloring constructed by the above algorithm has discrepancy at least $\frac{n}{2}$.*

Overview of ideas. Interestingly, one arrives at the counter-example of Theorem 0.59 by *also* using the MWU technique—somewhat reminiscent of the fact that the probabilistic method can be used to prove both upper and lower bounds. Basically, we will

1. assign, for each $S \in \mathcal{R}$, the weight

$$\omega(S) = \exp\left(\mathrm{disc}(S)\right),$$

2. construct sets so that assigning colors by following the above algorithm *increases* the total weight substantially at each iteration i, and

3. argue that at the end, by the pigeonhole principle, one set of $\mathcal{R}$ must have

> weight at least $\frac{1}{m}$-th of the total weight, which then gives a lower bound on the discrepancy of this set.

This works because the weight function is exponential, and so dividing by m causes only a $\log m$ additive loss. That is, the logarithm of the total weight essentially gives a lower bound on the discrepancy of the maximum set.

As with adversarial arguments, we will *incrementally* construct the counter-example set system $\mathcal{R}$ over the n iterations. This is possible since the greedy algorithm, at the i-th iteration, ignores the elements $v_{i+1}, \ldots, v_n$, and so we don't need to have the set system fully constructed at the very start. Specifically,

on receiving the element $v_i \in X$ in the i-th iteration, we will add v_i to the sets of $\mathcal{R}$ such that the greedy algorithm is forced to assign v_i a color that makes our total weight go up by a large amount—in fact, almost double.

Therefore, roughly speaking, at the end, the total weight is $\Omega(2^n)$, and so due to our choice of the weight function, one set of $\mathcal{R}$ must have discrepancy at least

$$\ln \frac{\Omega(2^n)}{m} = \Omega(n - \ln m).$$

Proof of Theorem 0.59. We adaptively constructing the sets as the algorithm proceeds with the n iterations. We will maintain the following invariant:

at all times, exactly half the sets of $\mathcal{R}$ will have positive discrepancy, and exactly half will have negative discrepancy.

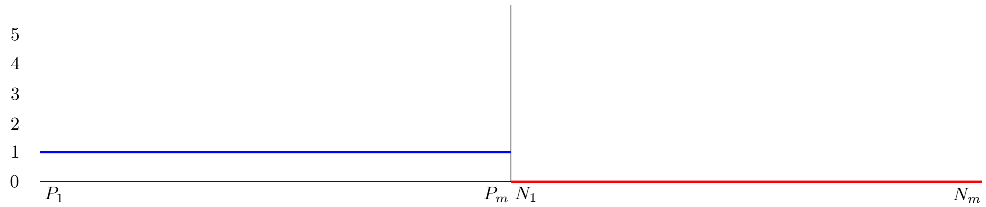
By relabeling, say that $\mathcal{R}$ consists of the $2m$ sets:

$$P_1, \ldots, P_m, \quad N_1, \ldots, N_m.$$

Initially, these $2m$ sets are empty and we now allocate vertices to them in the order $v_1, \ldots, v_n$.

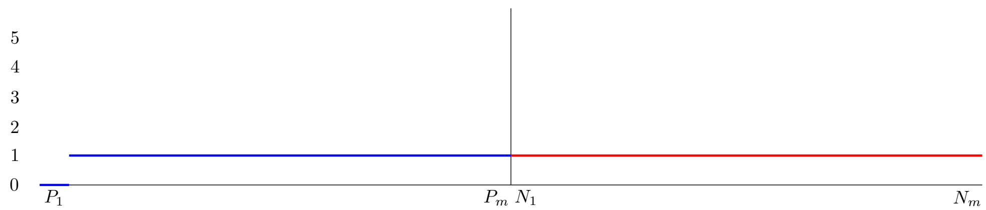
$\boldsymbol{v_1}$ **:** Add v_1 to $P_1, \ldots, P_m$. Set $\chi(v_1) = +1$.

$$\downarrow$$



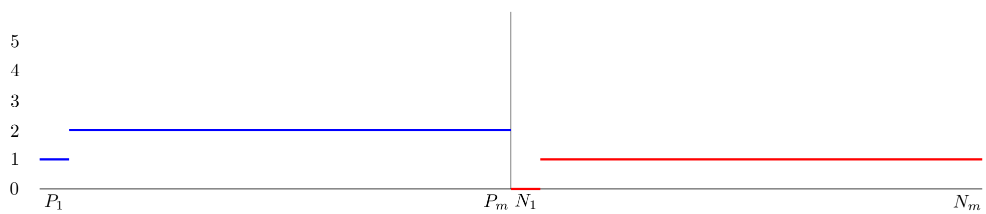
$\boldsymbol{v_2}$ **:** Add v_2 to P_1 and $N_1, \ldots, N_m$. To minimize $\mathrm{disc}(P_1)$, set $\chi(v_2) = -1$.

$\downarrow$



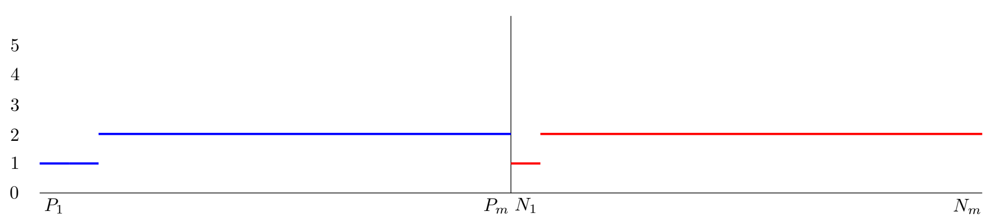
$\boldsymbol{v_3}$ **:** Add v_3 to N_1 and $P_1, \ldots, P_m$. To minimize $\mathrm{disc}(N_1)$, set $\chi(v_3) = +1$.

$\downarrow$



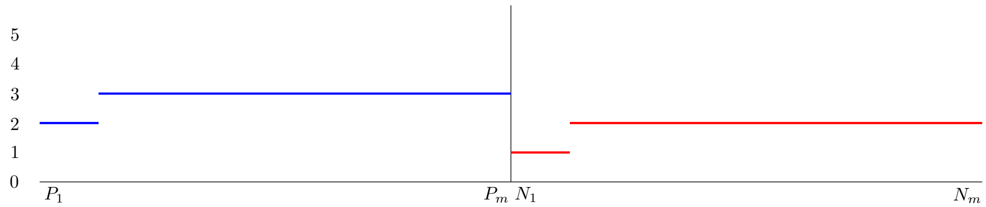
$\boldsymbol{v_4}$ **:** Add v_4 to P_2 and $N_1, \ldots, N_m$. To minimize $\mathrm{disc}(P_2)$, set $\chi(v_4) = -1$.

$\downarrow$



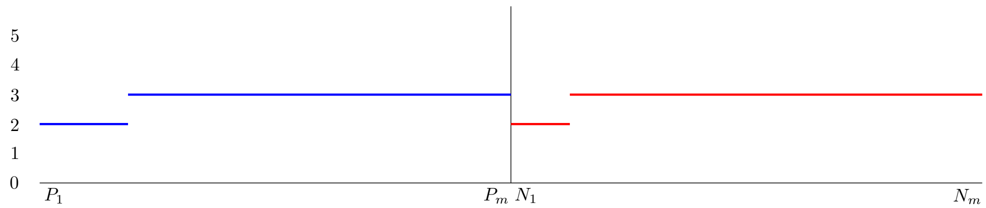
$\boldsymbol{v_5}$ **:** Add v_5 to N_2 and $P_1, \ldots, P_m$. To minimize $\mathrm{disc}(N_2)$, set $\chi(v_5) = +1$.

$\downarrow$



$\boldsymbol{v_6}$ **:** Add v_6 to P_3 and $N_1, \ldots, N_m$. To minimize $\mathrm{disc}(P_3)$, set $\chi(v_6) = -1$.

$$\downarrow$$



$$\vdots$$

$\boldsymbol{v_{2j}}$ **:** Add v_{2j} to P_j and $N_1, \ldots, N_m$. To minimize $\mathrm{disc}(P_j)$, set $\chi(v_{2j}) = -1$.

$\boldsymbol{v_{2j+1}}$ **:** Add v_{2j+1} to N_j and $P_1, \ldots, P_m$. To minimize $\mathrm{disc}(N_j)$, set $\chi(v_{2j+1}) = -1$.

At each iteration, we alternately increase the discrepancy of P_m or N_m by 1. Thus at the end, for $m \geq n$, we will have

$$\mathrm{disc}(P_m) = \mathrm{disc}(N_m) = \frac{n}{2}.$$

$\square$

Remark: If the $2m$ constructed sets are not distinct, we can always take an additional $\log m$ elements, and add a distinct subset of these elements to distinct sets of $\mathcal{R}$. This makes each set of $\mathcal{R}$ distinct, and can only change the discrepancy by an additive factor of $\log m$.

Remark: The construction is basically using a maximum discrepancy set as a 'lever' to increase the discrepancy of half the sets at each iteration:

alternately, we raise the discrepancy of all negative sets by using a positive discrepancy set, and raise the discrepancy of all positive sets by using a negative discrepancy set.

Bibliography and discussion. This example was constructed by the authors for pedagogical reasons. Similar constructions are well-known for other problems.

Discrepancy: Polynomial Function Bounds

Now, the question is whether an iterative coloring approach yields a coloring with the desired total discrepancy. More precisely, let us try the following approach.

We color the elements in order $v_1, \ldots, v_n$, guided by their weights:

> for a parameter $c \geq 1$ to be fixed later, set the weight of each $S \in \mathcal{R}$ as
>
> $$W(S) = \operatorname{disc}(S)^c,$$
>
> where $\operatorname{disc}(S)$ is the current discrepancy of S. Then the total weight is
>
> $$W(\mathcal{R}) = \sum_{S \in \mathcal{R}} W(S) = \sum_{S \in \mathcal{R}} \operatorname{disc}(S)^c.$$

Now, let W^{i+1} be the value of $W(\mathcal{R})$ after coloring v_i. At the start of the algorithm each S has discrepancy 0, and so $W^1 = 0$. In step i, we color element v_i in such a way that minimizes W^{i+1}. In particular,

> for a $S \in \mathcal{R}$ containing v_i and with $\operatorname{disc}(S) \neq 0$, the average weight of S after coloring v_i, over the two possible color choices, is
>
> $$\begin{aligned} \mathrm{E}\left[\left(\operatorname{disc}^{i+1}(S)\right)^c\right] &= \frac{1}{2}\left(\operatorname{disc}(S)+1\right)^c + \frac{1}{2}\left(\operatorname{disc}(S)-1\right)^c \\ &= \operatorname{disc}(S)^c \left(\frac{1}{2}\left(1+\frac{1}{\operatorname{disc}(S)}\right)^c + \frac{1}{2}\left(1-\frac{1}{\operatorname{disc}(S)}\right)^c\right) \\ &\leq \operatorname{disc}(S)^c \left(\frac{1}{2}\left(\exp\left(\frac{c}{\operatorname{disc}(S)}\right) + \exp\left(-\frac{c}{\operatorname{disc}(S)}\right)\right)\right) \end{aligned}$$
>
> Using Fact 0.52,
>
> $$\leq \operatorname{disc}(S)^c \cdot \exp\left(\frac{c^2}{2\operatorname{disc}(S)^2}\right).$$

We consider two cases, for a parameter a to be fixed later:

> **$\operatorname{disc}(S) < a$.** In this case, regardless of the color given to v_i, we have
>
> $$\operatorname{disc}^{i+1}(S) \leq a \qquad \Longrightarrow \qquad \left(\operatorname{disc}^{i+1}(S)\right)^c \leq a^c.$$

$\mathbf{disc}\,(\boldsymbol{S}) \geq \boldsymbol{a}$. Then we have

$$E\left[\left(\mathrm{disc}^{i+1}(S)\right)^c\right] \leq \left(\mathrm{disc}^i(S)\right)^c \cdot \exp\left(\frac{c^2}{2a^2}\right).$$

From the discussion so far, it follows that there is a choice of a color for v_i such that

$$\begin{aligned}
W^{i+1} &\leq \mathrm{E}\left[\sum_{S\in\mathcal{R}} W^{i+1}(S)\right] = \sum_{S\in\mathcal{R}} \mathrm{E}\left[W^{i+1}(S)\right] \\
&= \sum_{\substack{S\in\mathcal{R}:\\ \mathrm{disc}^i(S)<a}} \mathrm{E}\left[\left(\mathrm{disc}^{i+1}(S)\right)^c\right] + \sum_{\substack{S\in\mathcal{R}:\\ \mathrm{disc}^i(S)\geq a}} \left(\mathrm{disc}^{i+1}(S)\right)^c \\
&\leq \sum_{\substack{S\in\mathcal{R}:\\ \mathrm{disc}^i(S)<a}} a^c + \sum_{\substack{S\in\mathcal{R}:\\ \mathrm{disc}^i(S)\geq a}} \left(\mathrm{disc}^i(S)\right)^c \cdot \exp\left(\frac{c^2}{2a^2}\right) \\
&\leq m\,a^c + \exp\left(\frac{c^2}{2a^2}\right) \cdot W^i
\end{aligned}$$

Setting $a = \sqrt{nc}$,

$$\leq m(nc)^{\frac{c}{2}} + \exp\left(\frac{c}{2n}\right) \cdot W^i.$$

Using $W^1 = 0$ and unrolling the induction, we obtain

$$W^{n+1} = \sum_{i=0}^{n-1} \exp\left(\frac{c\,i}{2n}\right) \cdot m(nc)^{c/2} \leq n\exp\left(\frac{c}{2}\right) \cdot m(nc)^{c/2}.$$

Thus at the end of the algorithm, for each $S \in \mathcal{R}$, we have

$$\mathrm{disc}(S)^c \leq W^{n+1} \leq n\exp\left(\frac{c}{2}\right) m(nc)^{c/2},$$

implying that

$$\mathrm{disc}(S) \leq (nm)^{\frac{1}{c}} \exp\left(\frac{1}{2}\right) (nc)^{\frac{1}{2}}.$$

Setting $c = \ln nm$, we get the desired result—assuming that $m \geq n$.

Bibliography and discussion. The algorithm and its analysis was constructed here for pedagogical purposes.

Discrepancy: General Case

Given a set system $(X, \mathcal{R})$, a two-coloring of X, say denoted by χ, is an assigment of a '-1' or a '$+1$' color to each vertex of X. That is,

$$\chi\colon X \to \{-1, +1\}.$$

Then the *discrepancy* of χ with respect to $\mathcal{R}$, is defined as:

$$\operatorname{disc}_\chi(\mathcal{R}) = \max_{S\in\mathcal{R}} \left| \sum_{v\in X} \chi(v) \right|.$$

We give a MWU algorithm to compute a two-coloring with small discrepancy.

Theorem 0.50. *Let $(X, \mathcal{R})$ be a finite set system with $X = \{v_1, \ldots, v_n\}$ and $m = |\mathcal{R}|$. Then there is a deterministic MWU algorithm that computes a two-coloring of X with discrepancy $O\left(\sqrt{n \ln m}\right)$.*

Overview of ideas. We will color the elements of X sequentially, in the order $v_1, \ldots, v_n$, with a $+1$ or a -1 color. The elements that are so far uncolored will have color 0.

The idea is to maintain a weight for each set, where this weight depends *exponentially* on the current discrepancy of that set.

> Let $\eta > 0$ be a parameter to be set later.
>
> Define the weight of any $S \in \mathcal{R}$ as:
>
> $$W(S) = \exp\left(\eta \cdot \operatorname{disc}(S)\right),$$
>
> where $\operatorname{disc}(S)$ denotes the current discrepancy of S. That is, with the so-far uncolored elements having color 0.
>
> Set
>
> $$W(\mathcal{R}) = \sum_{S\in\mathcal{R}} W(S) = \sum_{S\in\mathcal{R}} \exp\left(\eta \cdot \operatorname{disc}(S)\right).$$

As with the MWU technique, when coloring element v_i, we will assign it a color that minimizes $W(\mathcal{R})$.

The key technical lemma is to show that, at each iteration, there is a choice of color for v_k such that the sum $W(\mathcal{R})$ grows slowly.

This then implies that no set can have too large a discrepancy.

Assume we have colored the elements $v_1, \ldots, v_{k-1}$ and now have to assign a color to v_k.

Let $\mathrm{disc}_k(\cdot)$ be the discrepancy and $W_k(\cdot)$ the weights, at the start of the k-th iteration. Note that for all $S \in \mathcal{R}$,

$$\mathrm{disc}_1(S) = 0 \qquad \text{and} \qquad W_1(S) = 1.$$

Claim 0.51. *At the start of the k-th iteration, let $\mathcal{R}' \subseteq \mathcal{R}$ be the sets with $\mathrm{disc}_k(\cdot) \neq 0$:*

$$\mathcal{R}' = \{S \in \mathcal{R} \colon \ \mathrm{disc}_k(S) \neq 0\}.$$

Then we can assign a color to v_k—that is, a $+1$ or a -1 value—such that

$$\sum_{S \in \mathcal{R}'} W_{k+1}(S) \leq \left(\frac{e^{\eta} + e^{-\eta}}{2}\right) \cdot \sum_{S \in \mathcal{R}'} W_k(S).$$

Proof. Set the color of v_k to $+1$ or -1 with equal probability.

If a $S \in \mathcal{R}$ does not contain v_k, its discrepancy does not change, and so for these sets, we have

$$W_{k+1}(S) = W_k(S) \leq \left(\frac{e^{\eta} + e^{-\eta}}{2}\right) W_k(S).$$

Otherwise, for any $S \in \mathcal{R}$ containing v_k and with $\mathrm{disc}(S) \neq 0$, the discrepancy of S increases by 1 or decreases by 1 with equal probability. Thus for any $S \in \mathcal{R}'$ containing v_k,

$$\begin{aligned} \mathrm{E}\left[W_{k+1}(S)\right] &= \frac{1}{2} \cdot e^{\eta(\mathrm{disc}_k(S)+1)} + \frac{1}{2} \cdot e^{\eta(\mathrm{disc}_k(S)-1)} \\ &= e^{\eta\, \mathrm{disc}_k(S)} \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right). \end{aligned}$$

By linearity of expectation, we have

$$\begin{aligned} \mathrm{E}\left[\sum_{S \in \mathcal{R}'} W_{k+1}(S)\right] &= \sum_{S \in \mathcal{R}'} \mathrm{E}\left[W_{k+1}(S)\right] \leq \sum_{S \in \mathcal{R}'} e^{\eta\, \mathrm{disc}_k(S)} \left(\frac{e^{\eta} + e^{-\eta}}{2}\right) \\ &= \left(\frac{e^{\eta} + e^{-\eta}}{2}\right) \cdot \sum_{S \in \mathcal{R}'} W_k(S). \end{aligned}$$

Thus for one of the two choices for the color of v_k, the desired statement holds.

□

Remark: The use of probability in the above proof is purely for 'implementing' an averaging argument. Essentially, we showed that

$$\underbrace{\sum_{S \in \mathcal{R}'} W\big(S \mid \ \mathrm{color}(v_k) = +1\big)}_{W_{k+1}(\mathcal{R}') \text{ assuming } \mathrm{color}(v_k) = +1} + \underbrace{\sum_{S \in \mathcal{R}'} W\big(S \mid \ \mathrm{color}(v_k) = -1\big)}_{W_{k+1}(\mathcal{R}') \text{ assuming } \mathrm{color}(v_k) = -1}$$

$$
\begin{aligned}
&= \sum_{S \in \mathcal{R}'} W\big(S \mid \operatorname{color}(v_k) = +1\big) + W\big(S \mid \operatorname{color}(v_k) = -1\big) \\
&\le \sum_{S \in \mathcal{R}'} e^{\eta} \cdot W_k(S) + e^{-\eta} \cdot W_k(S) \\
&= \big(e^{\eta} + e^{-\eta}\big) \cdot \sum_{S \in \mathcal{R}'} W_k(S),
\end{aligned}
$$

and so one of the two sums must be at most $\frac{1}{2}$ of the R.H.S. above.

For the moment, assume that for all $S \in \mathcal{R}$ and $k > 0$, we always have $\operatorname{disc}_k(S) \neq 0$. Then we're done:

Upper and lower bounding the total weight, we get

$$\max_{S \in \mathcal{R}} W_{n+1}(S) \;\le\; W_{n+1}(\mathcal{R}) \;\le\; W_1(\mathcal{R}) \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right)^n .$$

Using the inequality $e^{\eta} + e^{-\eta} \le 2e^{\eta^2/2}$ (Fact 0.52 below), and that $W_1(\mathcal{R}) = m$,

$$\exp\left(\eta \cdot \max_{S \in \mathcal{R}} \operatorname{disc}_{n+1}(S)\right) \;\le\; W_{n+1}(\mathcal{R}) \;\le\; m \cdot e^{n\eta^2/2}.$$

Taking logarithms,

$$\max_{S \in \mathcal{R}} \operatorname{disc}_{n+1}(S) \le \frac{\ln m}{\eta} + \frac{n\eta}{2}.$$

The above is minimized by setting $\eta = \Theta\left(\sqrt{\frac{\ln m}{n}}\right)$, giving the desired upper bound on the discrepancy for each set.

Fact 0.52. *For* $\eta \in \mathbb{R}$,

$$\frac{e^{\eta} + e^{-\eta}}{2} \le e^{\eta^2/2}.$$

Proof. Using Taylor series at 0 gives, for any $\eta > 0$,

$$
\begin{aligned}
e^{\eta} &= 1 + \frac{\eta}{1!} + \frac{\eta^2}{2!} + \frac{\eta^3}{3!} + \frac{\eta^4}{4!} + \cdots \\
e^{-\eta} &= 1 - \frac{\eta}{1!} + \frac{\eta^2}{2!} - \frac{\eta^3}{3!} + \frac{\eta^4}{4!} + \cdots
\end{aligned}
$$

Adding them up cancels the linear term—so the quadratic term becomes the dominant one for $\eta < 1$—and we get

$$e^{\eta} + e^{-\eta} = 2\left(1 + \frac{\eta^2}{2!} + \frac{\eta^4}{4!} + \frac{\eta^6}{6!} + \frac{\eta^8}{8!} + \cdots\right)$$

Using the fact that $(2i)! \geq 2^i i!$,

$$\begin{aligned} &< 2\left(1 + \frac{(\eta^2)}{2^1 \cdot 1!} + \frac{(\eta^2)^2}{2^2 \cdot 2!} + \frac{(\eta^2)^3}{2^3 \cdot 3!} + \frac{(\eta^2)^4}{2^4 \cdot 4!} + \cdots\right) \\ &= 2\left(1 + \frac{(\eta^2/2)}{1!} + \frac{(\eta^2/2)^2}{2!} + \frac{(\eta^2/2)^3}{3!} + \frac{(\eta^2/2)^4}{4!} + \cdots\right) \\ &= 2\, e^{\eta^2/2}. \end{aligned}$$

□

The above does not *quite* work—we used Claim 0.51 which only applies to sets of $\mathcal{R}$ with $\mathrm{disc}_k(\cdot) \neq 0$. Indeed, the restriction to sets with $\mathrm{disc}_k(\cdot) \neq 0$ is necessary for Claim 0.51 to be correct:

> The key property in Claim 0.51 is that the discrepancy of each $S \in \mathcal{R}'$ can *both* increase or decrease by 1. This is what allows us to upper bound the *average* multiplicative factor increase in the weight of each $S \in \mathcal{R}'$ by $(e^{\eta} + e^{-\eta})/2$.
>
> However, this is not true when $\mathrm{disc}(S) = 0$—then the discrepancy of S can only increase by 1, no matter what color is given to v_k, and so the multiplicative factor becomes e^{η}. This is too big—by a factor of roughly 2 at each iteration, and so with a 2^n factor at the end that gives a useless bound.

We now present two ways to get around this problem:

Bounding total increase in weights. The weight function is the same as earlier:

$$W(S) = \exp(\eta \cdot \mathrm{disc}(S)).$$

As before:

1. we choose the color of v_k by considering $S \in \mathcal{R}$ with $\mathrm{disc}(S) > 0$, and then applying Claim 0.51.
2. The total weight of the sets with $\mathrm{disc}(S) > 0$ increases by a multiplicative factor of at most $\left(\frac{e^{\eta}+e^{-\eta}}{2}\right)$.

However now, additionally, the weight of each set with $\mathrm{disc}(S) = 0$, goes from 1 to e^{η}. But this is not really a problem:

> the weight of S is already small when $\mathrm{disc}(S) = 0$—it is $e^{\eta \cdot 0} = 1$, and will become e^{η}. This is small-enough to be incorporated in the calculation without significantly changing the upper bound.

Taking both types of weight changes into account, we have

$$W_{k+1}(\mathcal{R}) \leq me^{\eta} + W_k \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right)$$

Opening it up inductively,

$$\begin{aligned}
W_{n+1}(\mathcal{R}) &\leq me^{\eta} + \left(m \cdot e^{\eta} + W_{n-1} \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right)\right) \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right) \\
&= me^{\eta} + m \cdot e^{\eta} \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right) + W_{n-1} \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right)^2 \\
&\vdots \\
&\leq \left(\sum_{i=0}^{n-1} me^{\eta} \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right)^i\right) + m \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right)^n \\
&\leq m\, n\, e^{\eta} \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right)^n + m \cdot \left(\frac{e^{\eta} + e^{-\eta}}{2}\right)^n \\
&\leq 2\, m\, n\, e^{\eta} \left(\frac{e^{\eta} + e^{-\eta}}{2}\right)^n .
\end{aligned}$$

Now the previous double-counting argument finishes the proof as before:

$$\exp\left(\eta \cdot \max_{S \in \mathcal{R}} \mathrm{disc}_{n+1}(S)\right) \;\leq\; W_{n+1}(\mathcal{R}) \;\leq 2m\, n\, e^{\eta}\, e^{n\eta^2/2}.$$

Taking logarithms,

$$\max_{S \in \mathcal{R}} \mathrm{disc}_n(S) = O\left(\frac{\ln mn}{\eta} + n\eta\right).$$

Setting $\eta = \Theta\left(\sqrt{\frac{\ln mn}{n}}\right)$ gives an upper bound of $O\left(\sqrt{n \ln m}\right)$, assuming $m \geq n$.

Using a different weight function. The trick here—on seeing the multiplicative factor of $\left(\frac{e^{\eta}+e^{-\eta}}{2}\right)$—is to slightly modify the weight function so that even when $\mathrm{disc}(S) = 0$, the weight increases by a smaller multiplicative factor.

We set the new weight function, denoted by $\omega(\cdot)$, to be:

$$\omega(S) = \frac{\exp(\eta \cdot \mathrm{disc}(S)) + \exp(-\eta \cdot \mathrm{disc}(S))}{2}. \tag{0.53}$$

Now note that even when $\mathrm{disc}_k(S) = 0$ with $\omega_k(S) = 1$, we have

$$\omega_{k+1}(S) = \frac{\exp(\eta) + \exp(-\eta)}{2},$$

which is the precise multiplicative increase we wanted.

Further, the general upper bound on the multiplicative weight increase continues to hold, as before, for the case $\mathrm{disc}(S) \neq 0$:

$$\begin{aligned}
\mathrm{E}\left[\omega_{k+1}(S)\right] &= \frac{1}{2}\left(\frac{e^{\eta\cdot(\mathrm{disc}_k(S)+1)} + e^{-\eta\cdot(\mathrm{disc}_k(S)+1)}}{2}\right) + \frac{1}{2}\left(\frac{e^{\eta\cdot(\mathrm{disc}_k(S)-1)} + e^{-\eta\cdot(\mathrm{disc}_k(S)-1)}}{2}\right) \\
&= \frac{e^{\eta\,\mathrm{disc}_k(S)}\cdot e^{\eta}}{4} + \frac{e^{-\eta\,\mathrm{disc}_k(S)}\cdot e^{-\eta}}{4} + \frac{e^{\eta\,\mathrm{disc}_k(S)}\cdot e^{-\eta}}{4} + \frac{e^{-\eta\,\mathrm{disc}_k(S)}\cdot e^{\eta}}{4} \\
&= \frac{e^{\eta\,\mathrm{disc}_k(S)}}{2}\left(\frac{e^{\eta}+e^{-\eta}}{2}\right) + \frac{e^{-\eta\,\mathrm{disc}_k(S)}}{2}\left(\frac{e^{\eta}+e^{-\eta}}{2}\right) \\
&= \left(\frac{e^{\eta\,\mathrm{disc}_k(S)} + e^{-\eta\,\mathrm{disc}_k(S)}}{2}\right)\cdot\left(\frac{e^{\eta}+e^{-\eta}}{2}\right) \\
&= \omega_k(S)\cdot\left(\frac{e^{\eta}+e^{-\eta}}{2}\right).
\end{aligned}$$

Now the previous double-counting argument finishes the proof.

Remark: Here is one way to naturally derive the weight function given in Equation (0.53).

Our goal is to minimize $\mathrm{disc}(S)$—in other words, for each $S \in \mathcal{R}$, the number of '$+1$' colors should not be too large, and neither should the number of '-1' colors.

Our earlier weight function, $\exp(\eta\,\mathrm{disc}(S))$, was capturing this compactly using the absolute value function. But the drawback of this is that it made it insensitive to the case when $\mathrm{disc}(S) = 0$.

We can fix this by *separately* adding the two exponential constraints—one prohibiting too many '$+1$' colors, and the other prohibiting too many '-1' colors:

For each $S \in \mathcal{R}$, let P_S be the number of elements of color '$+1$', and N_S the number of elements of color '-1'.

Then we minimize the weight function

$$\exp\big(\eta\,(P_S - N_S)\big) + \exp\,\big(\eta(N_S - P_S)\big).$$

This is exactly Equation (0.53) scaled by a factor of 2! The constant 2 is not important and could have been omitted—the calculation without it gives the same bound.

Bibliography and discussion. The hyperbolic cosine algorithm is from [Bec81; BF81].
Another way one can arrive at the function $\frac{1}{2}\,(e^{\eta} + e^{-\eta})$ is via the proof of the tail bound used to prove the $O\left(\sqrt{n \ln m}\right)$ bound for discrepancy via a random coloring; see [You95].

[BF81] József Beck and Tibor Fiala. ""Integer-making" theorems". In: *Discrete Applied Mathematics* 3.1 (1981), pp. 1–8.

[Bec81] József Beck. "Van der waerden and Ramsey type games". In: *Combinatorica* 1.2 (1981), pp. 103–116.

[You95] N. E. Young. "Randomized Rounding Without Solving the Linear Program". In: *Proceedings of the Sixth Annual Symposium on Discrete Algorithms (SODA)*. 1995, pp. 170–178.