Asymptotic Complexity of Flood-It under a Random Policy

Daan Noordenbos

December 4, 2023

Abstract

The asymptotic complexity of the one player board game Flood-It is studied under a random policy. It is shown that for an n by m board with c colors that under a random policy, the expected amount of steps is of asymptotic order $O(n(1-c)+(m-1)(c-1))$.

1 Introduction

The board game Flood-It is played on an n by m rectangular board with up to c colors. The objective is to fill the entire board with a single color, starting from the top-left corner. On each turn, a color is chosen, and the region which has the same color as and is connected to the top-left corner has its colored changed to the chosen color. Figure 1 gives an example of game play.

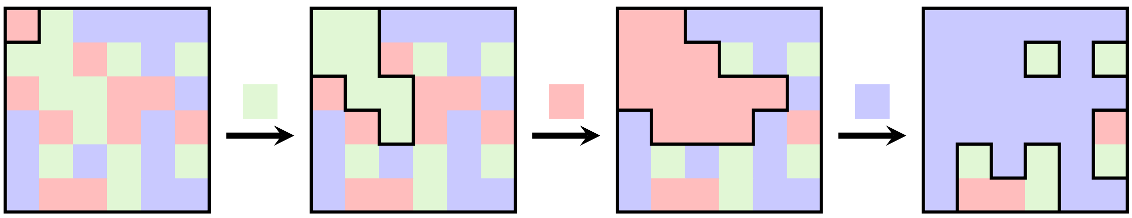

Figure 1: Example game play of Flood-it on a 6×6 board with three colors.

The goal of the game is to minimize the number of turns to fill the board. Clifford et al. (2012) have shown that determining the minimum amount of steps required to flood the board is NP-hard if $c \geq 3$. This paper will consider the performance of a random policy and give the asymptotic order of the expected amount of steps to flood the board.

1.1 Notation and definitions

The quantity of interest is the amount of steps to flood a board which is n squares wide and m squares tall with at most c colors. This quantity, the flooding time, will be denoted by $T_{n\times m}$ or T_n when $m=1$ (c is taken to be known). The term, the flooded region, will refer the region which has the same color as and is connected to the top-left corner. Lastly, $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$ is denoted by $f(x) \overset{\mathrm{L}}{\sim} g(x)$.

2 Distribution of T_n

To start an n by 1 board is considered. For such a board progress can only be made by picking the color which neighbours the flooded region. This happens with probability $\frac{1}{c}$, and if this happens there is a $\frac{1}{c}$ probability that the next square is of the same color. In fact, each turn the amount of squares that are added to the flooded region follows a geometric distribution with $p = 1 - \frac{1}{c}$. It is important to note that initially, the size of the flooded region is not zero, but one plus the geometric random variable. Consequently, if S_k denotes the size of the flooded region after k turns then,

$$S_k = 1 + G_0 + \sum_{i=1}^{k} G_i,$$

where G_i are i.i.d. and have a Geom $\left(1 - \frac{1}{c}\right)$ distribution. Using this T_n can be characterized as follows,

$$T_n = \inf\{k \in \mathbb{Z}_+ : S_k \geq n\} = \inf\{k \in \mathbb{Z}_+ : \sum_{i=0}^{k} G_i \geq n - 1\}.$$

Since each G_i is non-negative, it follows that,

$$\begin{aligned}
\mathbb{P}(T_n \leq k) &= \mathbb{P}\left(\sum_{i=0}^{k} G_i \geq n-1\right) && \text{(Non-negativity of } G_i\text{)} \\
&= 1 - \sum_{j=1}^{n-2} \mathbb{P}\left(\sum_{i=0}^{k} G_i = i\right) \\
&= 1 - \sum_{j=1}^{n-2} \binom{j+k}{j} (1-p)^j p^{k+1} && \text{(Sum geometric is negative binomial)} \\
&= 1 - I_p(k+1, n-1) && \text{(Definition incomplete beta function)} \\
&= I_{1-p}(n-1, k+1) && \text{(Property incomplete beta function).}
\end{aligned}$$

The final quantity is the CDF of a negative binomial distribution with parameters $n-1$ and $\frac{1}{c}$. Therefore, $T_n \sim \mathrm{NB}\left(n-1, \frac{1}{c}\right)$.

3 Asymptotics of $\mathbb{E}[T_{n \times m}]$

The distribution of $T_{n\times m}$ cannot be determined, but $\mathbb{E}[T_{n\times m}]$ can be bounded. To achieve this, the columns of the board are decoupled in a manner that increases the expected number of steps to flood the board. The specific decoupling can be seen in Figure 2. The flooding time of a decoupled $n \times m$ board is denoted by $\tilde{T}_{n\times m}$. The decoupling of the columns removes the possibility of the squares changing color. As a result, the distribution of the amount of sequential squares which have the same color in each column has a lower stochastic order than in the original board after any amount of moves have been made. This is because the color switching can only increase the length of sequential strings of squares of the same color. Having shorter sequences of squares of the same color increase the number of steps needed to flood a column. Therefore, $\tilde{T}_{n\times m}$ has a greater stochastic order than $T_{n\times m}$, i.e., it takes longer to flood the decoupled board.

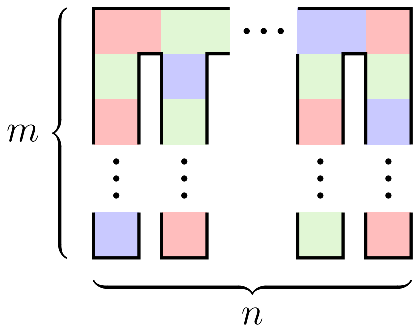

Figure 2: An $n \times m$ board with decoupled columns.

Bounding the expectation of $\tilde{T}_{n\times m}$ is straightforward. The flooding of the decoupled board occurs in two stages: First, the top row must be flooded, and then each column must be flooded. This is where the advantage of decoupling comes in, the time to flood each column is now independent of each other. The top row takes T_n steps to flood, and in the worst case scenario, each of the n columns take T_m steps to flood. Note that, unlike in the case of an $n \times 1$ board, the first square is not free. Therefore, each column takes T_m steps and not T_{m-1}. As a result, the total flooding time has a smaller stochastic order than,

$$T_n + \max\{T_{m,1}, \dots, T_{m,n}\}.$$

To continue an auxiliary Lemma and Theorem 3.2 are required.

Lemma 3.1. *For n i.i.d. random variables X_i and $t \geq 0$:*

$$\mathbb{E}\left[\max\{X_1, \dots, X_n\}\right] \leq \frac{1}{t}\log(nM_X(t))$$

Proof. Consider the following application of Jensen's inequality:

$$\exp\left(t\mathbb{E}\left[\max_{i\in\{1,\dots,n\}}\{X_i\}\right]\right) \leq \mathbb{E}\left[\exp(t\max_{i\in\{1,\dots,n\}}\{X_i\})\right] \leq \mathbb{E}\left[\sum_{i=1}^{n}\exp(tX_i)\right] = nM_X(t).$$

Applying a logarithm, a monotonic transformation, and subsequently dividing my a positive number does not affect the inequality sign. Therefore, $\mathbb{E}\left[\max\{X_1, \dots, X_n\}\right] \leq \frac{1}{t}\log(nM_X(t))$. $\square$

Theorem 3.2. *Let X_i be n i.i.d. random variables with distribution* $\mathrm{NB}(m,p)$*. If $n \to \infty$ and $m \to \infty$ such that $m \overset{\mathrm{L}}{\sim} n^\alpha$ for $\alpha > 0$ then $O(\mathbb{E}[X_{(n)}]) = O\left(m\frac{1-p}{p}\right)$.*

Proof. By Lemma 3.1, $\mathbb{E}[X_{(n)}] \leq \frac{1}{t}\log\left(nM_X(t)\right)$ for all $t \geq 0$. To obtain a suitable bound a t value is chosen which minimizes the right-hand side. Directly minimizing does not yield a closed form of t, so instead, a suitable t is chosen by minimizing the normal approximation of X. The normal approximation in question, $\mathcal{N}\left(m\frac{1-p}{p}, m\frac{1-p}{p^2}\right)$, produces the following choice of t:

$$\min_{t\geq 0}\frac{1}{t}\log\left(n\exp\left(m\frac{1-p}{p}t + \frac{1}{2}m\frac{1-p}{p^2}t^2\right)\right) = \min_{t\geq 0}\frac{\log n}{t} + m\frac{1-p}{p} + \frac{1}{2}m\frac{1-p}{p^2}t$$

$$\implies t^* = \sqrt{\frac{2p^2\log n}{m(1-p)}}.$$

With this choice of t an asymptotic analysis of the bound can be performed:

$$\begin{aligned}\frac{1}{t^*}\log\left(nM_X(t)\right) &= \frac{\log n}{t^*} + \frac{m}{t^*}\log\left(\frac{p}{1-(1-p)e^{t^*}}\right)\\ &= \underbrace{\sqrt{\frac{m(1-p)\log n}{2p^2}}}_{\text{Asymptotically irrelevant}} + m\underbrace{\frac{1}{t^*}\log\left(\frac{p}{1-(1-p)e^{t^*}}\right)}_{\to\frac{1-p}{p},\ n,m\to\infty \implies t^*\to 0 \text{ apply L'H}}\\ &\overset{\mathrm{L}}{\sim} m\left(\frac{1-p}{p}\right).\end{aligned}$$

Note that $t^* \to 0$ because $m \overset{\mathrm{L}}{\sim} n^\alpha$. From the asymptotics of the bound it is evident that $O(\mathbb{E}[X_{(n)}]) = O\left(m\frac{1-p}{p}\right)$. □

Theorem 3.2 informs us that if $n \to \infty$ and $m \to \infty$ such that $m \overset{\mathrm{L}}{\sim} n^\alpha$ for $\alpha > 0$ then,

$$O\left(\mathbb{E}\left[T_n + \max\{T_{m,1}, \ldots, T_{m,n}\}\right]\right) = O(n(c-1) + (m-1)(c-1)).$$

$T_n + \max\{T_{m,1}, \ldots, T_{m,n}\}$ has a greater stochastic order than $\tilde{T}_{n\times m}$ which in turn has a greater stochastic order $T_{n\times m}$, therefore it can be concluded that $O(\mathbb{E}[T_{n\times m}]) = O(n(1-c) + (m-1)(c-1))$.

References

Clifford, Raphaël, Markus Jalsenius, Ashley Montanaro, and Benjamin Sach (2012). The complexity of flood filling games. *Theory of Computing Systems 50*(1), 72–92.