Extending Wormald's Differential Equation Method to One-sided Bounds

Patrick Bennett[*] Calum MacRury [†]

Abstract

In this note, we formulate a "one-sided" version of Wormald's differential equation method. In the standard "two-sided" method, one is given a family of random variables which evolve over time and which satisfy some conditions including a tight estimate of the expected change in each variable over one time step. These estimates for the expected one-step changes suggest that the variables ought to be close to the solution of a certain system of differential equations, and the standard method concludes that this is indeed the case. We give a result for the case where instead of a tight estimate for each variable's expected one-step change, we have only an upper bound. Our proof is very simple, and is flexible enough that if we instead assume tight estimates on the variables, then we recover the conclusion of the standard differential equation method.

1 Introduction

In the most basic setup of Wormald's differential equation method, one is given a sequence of random variables $(Y(i))_{i=0}^{\infty}$ derived from some random process which evolves step by step. The random variables $(Y(i))_{i=0}^{\infty}$ all implicitly depend on some $n \in \mathbb{N}$, and the goal is understand their typical behaviour as $n \to \infty$.

Our running example is based on the the Erdős–Rényi **random graph process** $(G_i)_{i=0}^{m}$ on vertex set $[n] := \{1, \dots, n\}$ where $G_i = ([n], E_i)$ and $m, n \in \mathbb{N}$. Here $G_0 = ([n], E_0)$ is the empty graph, and G_{i+1} is constructed from G_i by drawing an edge e_{i+1} from $\binom{[n]}{2} \setminus E_i$ uniformly at random (u.a.r.), and setting $E_{i+1} := E_i \cup \{e_{i+1}\}$. Suppose that we wish to understand the size of the matching produced by the greedy algorithm as it executes on $(G_i)_{i=0}^{m}$. More specifically, when e_{i+1} arrives, the greedy algorithm adds e_{i+1} to the current matching if the endpoints of e_{i+1} were not previously matched. We will let $m = cn$, i.e. we will add a linear number of random edges. Observe that if $Y(i)$ is the number of edges of G_i matched by the algorithm, then $Y(i)$ is a function of $e_1, \dots, e_i$ (formally, $Y(i)$ is $\mathcal{H}_i$-measurable where $\mathcal{H}_i$ is the sigma-algebra generated from $e_1, \dots, e_i$). Then for $i < m$,

$$\mathbb{E}[\Delta Y(i) \mid \mathcal{H}_i] = \frac{\binom{n-2Y(i)}{2}}{\binom{n}{2} - i} = \left(1 - \frac{2Y(i)}{n}\right)^2 + O\left(\frac{1}{n}\right), \tag{1}$$

[*]Department of Mathematics, Western Michigan University (`patrick.bennett@wmich.edu`). Supported in part by Simons Foundation Grant #426894.

[†]Department of Computer Science, University of Toronto, Toronto, ON, Canada `cmacrury@cs.toronto.edu`

where $\Delta Y(i) := Y(i+1) - Y(i)$, and the asymptotics are as $n \to \infty$ (which will be the case for the remainder of this note). By scaling $(Y(i))_{i=0}^{m}$ by n, we can interpret the left-hand side of (1) as the "derivative" of $Y(i)/n$ evaluated at i/n. This suggests the following *differential equation*:

$$y'(t) = (1 - 2y(t))^2, \qquad y(0) = 0 \tag{2}$$

with initial condition $y(0) = 0$. Wormald's differential equation method allows us to conclude that **with high probability** (i.e. with probability tending to 1 as $n \to \infty$, henceforth abbreviated w.h.p.),

$$Y(m) = (1 + o(1))y(m/n), \tag{3}$$

where $y(t) := t/(1 + 2t)$ is the unique solution to (2).

Returning to the general setup of the differential equation method, suppose we are given a sequence of random variables $(Y(i))_{i=0}^{\infty}$ which implicitly depend on $n \in \mathbb{N}$. Assume that the one-step changes are bounded, i.e., there exists a constant $\beta \geq 0$ such that $|\Delta Y(i)| \leq \beta$ for each $i \geq 0$. Moreover, suppose each $Y(i)$ is determined by some sigma-algebra $\mathcal{H}_i$, and its *expected* one-step changes are described by some Lipshitz function $f = f(t, y)$. That is, for each $i \geq 0$,

$$\mathbb{E}[\Delta Y(i) \mid \mathcal{H}_i] = f(i/n, Y(i)/n) + o(1). \tag{4}$$

If $Y(0) = (1 + o(1))\widetilde{y}n$ for some constant $\widetilde{y}$, and $m = m(n)$ is not too large, then the differential equation method allows us to conclude that w.h.p. $Y(m)/n = (1 + o(1))y(m/n)$ for y which satisfies the differential equation suggested by (4), i.e.

$$y'(t) = f(t, y(t))$$

with initial condition $y(0) = \widetilde{y}$. In this note, we consider the case when we have an inequality in place of (4). We are motivated by applications to online algorithms in which one wishes to upper bound the performance of *any* online algorithm, opposed to just a *particular* algorithm. (See Section 1.1 for an example pertaining to online matching in $(G_i)_{i=0}^{m}$ as well as some discussion of further applications). We are also motivated by the existence of deterministic results of which we wanted to prove a random analogue. For example, consider the following classical result due to Petrovitch [9]:

Theorem 1. *Suppose $f : \mathbb{R}^2 \to \mathbb{R}$ is Lipschitz continuous, and $y = y(t)$ satisfies*

$$y'(t) = f(t, y(t)), \qquad y(c) = y_0.$$

Suppose $z = z(t)$ is differentiable and satisfies

$$z'(t) \leq f(t, z(t)), \qquad z(c) = z_0 \leq y_0.$$

Then $z(t) \leq y(t)$ for all $t \geq c$.

With the above result in mind (as well as the standard differential equation method), it's natural to wonder what can be said about a sequence of random variables $(Z_i)_{i=1}^{\infty}$ satisfying

$$\mathbb{E}[\Delta Z_i \mid \mathcal{H}_i] \leq f(i/n, Z_i/n) \tag{5}$$

instead of the equality version (4). More precisely, if $(Y_i)_{i=1}^{\infty}$ satisfies (4) and $Z_0 < Y_0$ then should it not follow that we likely have $Z_i \leq Y_i$ (perhaps modulo some relatively small error term) for all $i \geq 0$?

We briefly point out that if f is nonincreasing in its second variable, then the problem described in the previous paragraph is much easier. Indeed, whenever the random variable satisfies $Z_i - Y_i \leq 0$, it is also a supermartingale. More precisely, when $Z_i \leq Y_i$ we have that

$$\mathbb{E}[\Delta(Z_i - Y_i) \mid \mathcal{H}_i] \leq f(i/n, Z_i/n) - f(i/n, Y_i/n) \leq 0$$

by the monotonicity assumption. In this case, assuming the initial gap $|Z_0 - Y_0|$ is large enough, standard martingale techniques can be used to bound the probability that the supermartingale $Z_i - Y_i$ becomes positive. However, we would like to handle applications where we do not have this monotonicity assumption. For instance, in our running example, $f(t, z) = (1 - 2z)^2$ is *not* increasing in z.

Of course, the differential equation method in general deals with systems of random variables (and the associated systems of differential equations). So what can be said about systems of deterministic functions whose derivatives satisfy inequalities instead? It turns out that to generalize Theorem 1 to a system, we need the functions to be **cooperative**. We say the functions $f_j := \mathbb{R}^{a+1} \to \mathbb{R}, 1 \leq j \leq a$ are cooperative (respectively, **competitive**) if each f_j is nondecreasing (respectively, nonincreasing) in all of its $a+1$ inputs except for possibly the first input and the $(j+1)^{th}$ one. In other words, $f_j(t, y_1, \ldots y_a)$ is nondecreasing in all variables except possibly t and y_j. Note that some sources refer to a system with the cooperative property as being **quasimonotonic**. Observe that in the one-dimensional case $a = 1$, every function is cooperative/cooperate. The following theorem is folklore (see [11] for some relevant background, and Section 3 for a proof):

Theorem 2. *Suppose $f_j : \mathbb{R}^{a+1} \to \mathbb{R}, 1 \leq j \leq a$ are Lipschitz continuous and cooperative, and y_j satisfies*

$$y_j'(t) = f_j(t, y_1(t), \ldots, y_a(t)), \qquad 1 \leq j \leq a, \quad t \geq c. \tag{6}$$

Suppose $z_j, 1 \leq j \leq a$ are differentiable and satisfy $z_j(c) \leq y_j(c)$ and

$$z_j'(t) \leq f_j(t, z_1(t), \ldots, z_a(t)), \qquad 1 \leq j \leq a, \quad t \geq c. \tag{7}$$

Then $z_j(t) \leq y_j(t)$ for all $1 \leq j \leq a$, $t \geq c$.

Cooperativity is necessary in the sense that if we do not have it, then one can choose initial conditions for the functions y_j, z_j to make the conclusion of Theorem 2 fail. Indeed, suppose we do not have cooperativity, i.e. there exist j, j' with $j' \neq j+1$ and some points $\mathbf{p}, \mathbf{p}' \in \mathbb{R}^{a+1}$ that agree everywhere except for their j'^{th} coordinate, where we have $p_{j'} > p'_{j'}$, and $f_j(\mathbf{p}) < f_j(\mathbf{p}')$. Consider the following initial conditions:

$$\Big(c, y_1(c), \ldots, y_a(c)\Big) = \mathbf{p}, \qquad \Big(c, z_1(c), \ldots, z_a(c)\Big) = \mathbf{p}'.$$

Then we have that $z_j(c) = y_j(c) = p_{j+1} = p'_{j+1}$. Furthermore, $z_j'(c)$ could be as large as $f_j(\mathbf{p}') > f_j(\mathbf{p}) = y_j'(c)$ in which case clearly $z_j(t) > y_j(t)$ for some $t > c$.

Our main theorem in this paper, Theorem 3, is essentially the random analogue of Theorem 2. Before providing its formal statement, we expand upon why it is useful for proving *impossibility/hardness* results for online algorithms. The reader can safely skip Section 1.1 if they would first like to instead read Theorem 3.

1.1 Motivating Applications

The example considered in this section is closely related to the 1/2-impossibility (or hardness) result for an online stochastic matching problem considered by the second author, Ma and Grammel in [8]. In fact, in the latest arXiv version of [8], Theorem 3 is used explicitly to simplify the most technical step of the argument. Our theorem can also be used to simplify the proofs of the $\frac{1}{2}(1+e^{-2})$-impossibility result of Fu et al. (Theorem 2 in [6]), and the $1-\ln(2-1/e)$-impossibility result of Fata et al. (Lemma 5 in [3]). All of the aforementioned papers prove impossibility results for various online stochastic optimization problems – more specifically, hardness results for *online contention resolution schemes* [4] or *prophet inequalities* against an "ex-ante relaxation" [7]. We think that Theorem 3 will find further applications as a technical tool in this area.

Let us now return to the definition of the Erdős–Rényi random graph process $(G_i)_{i=0}^m$ as discussed in Section 1, where we again assume that $m = cn$ for some constant $c > 0$. Recall that (3) says that if $Y(m)$ is the size of the matching constructed by the greedy matching algorithm when executed on $(G_i)_{i=0}^m$, then w.h.p. $Y(m)/n = (1+o(1))y(c)$ where $y(c) = c/(1+2c)$. In fact, (3) can be made to hold with probability $1-o(1/n^2)$, and so $\mathbb{E}[Y(m)]/n = (1+o(1))c/(1+2c)$ after taking expectations.

The greedy matching algorithm is an example of an **online (matching) algorithm** on $(G_i)_{i=0}^m$. An online algorithm begins with the empty matching on G_0, and its goal is to build a matching of G_m. While it knows the distribution of $(G_i)_{i=0}^m$ upfront, it learns the *instantiations* of the edges sequentially and must execute online. Formally, in each step $i \geq 1$, it is presented e_i and then makes an irrevocable decision as to whether or not to include e_i in its current matching, based upon $e_1, \ldots, e_{i-1}$ and its previous matching decisions. Its output is the matching M_m, and its goal is to maximize $\mathbb{E}[|M_m|]$. Here the expectation is over $(G_i)_{i=1}^m$ and any randomized decisions made by the algorithm.

Suppose that we wish to prove that the greedy algorithm is asymptotically optimal. That is, for *any* online algorithm, if M_m is the matching it outputs on G_m, then $\mathbb{E}[|M_m|] \leq (1+o(1))\mathbb{E}[Y(m)]$. In order to prove this directly, one must compare the performance of any online algorithm to the greedy algorithm. This is inconvenient to argue, as there exist rare instantiations of $(G_i)_{i=0}^m$ in which being greedy is clearly sub-optimal.

We instead upper bound the performance of any online algorithm by $(1+o(1))y(c)n$. Let $(M_i)_{i=0}^m$ be the sequence of matchings constructed by an *arbitrary* online algorithm while executing on $(G_i)_{i=0}^m$. For simplicity, assume that the algorithm is deterministic so that M_i is $\mathcal{H}_i$-measurable. In this case, we can replace (1) with inequality. I.e., if $Z(i) := |M_i|$, then for $i < m$,

$$\mathbb{E}[\Delta Z(i) \mid \mathcal{H}_i] \leq \left(1 - \frac{2Z(i)}{n}\right)^2 + O\left(\frac{1}{n}\right). \tag{8}$$

Recall now the intuition behind the differential equation method. If we scale $(Z(i))_{i=0}^n$ by n, then we can interpret the left-hand side of (8) as the "derivative" of $Z(i)/n$ evaluated at i/n. This suggests the following differential inequality:

$$z' \leq (1-2z)^2, \tag{9}$$

with inital condition $z(0) = 0$. By applying Theorem 3 to $(Z(i))_{i=0}^m$, we get that

$$Z(m)/n \leq (1+o(1))y(c) \tag{10}$$

with probability $1-o(n^{-2})$. As a result, $\mathbb{E}[Z(m)] \leq (1+o(1))y(c)n$, and so we can conclude that greedy is asymptotically optimal.

2 Main Theorem

For any sequence $(Z(i))_{i=0}^{\infty}$ of random variables and $i \geq 0$, we will use the notation $\Delta Z(i) := Z(i+1) - Z(i)$. Note that given a **filtration** $(\mathcal{H}_i)_{i=0}^{\infty}$ (i.e., a sequence of increasing σ-algebras), we say that $(Z_j(i))_{i=0}^{\infty}$ is **adapted** to $(\mathcal{H}_i)_{i=0}^{\infty}$, provided Z_i is $\mathcal{H}_i$-measurable for each $i \geq 0$. Finally, we say that a stopping time I is **adapted** to $(\mathcal{H}_i)_{i=0}^{\infty}$, provided the event $\{I = i\}$ is $\mathcal{H}_i$-measurable for each $i \geq 0$.

Given $a \in \mathbb{N}$, suppose that $\mathcal{D} \subseteq \mathbb{R}^{a+1}$ is a bounded domain, and for $1 \leq j \leq a$, let $f_j : \mathcal{D} \to \mathbb{R}$. We assume that the following hold for each j:

a) f_j is L-Lipschitz,

b) $|f_j| \leq B$ on $\mathcal{D}$, and

c) the $(f_j)_{j=1}^{a}$ are cooperative.

Given $(0, \tilde{y}_1, \ldots, \tilde{y}_a) \in \mathcal{D}$, assume that $y_1(t), \ldots, y_a(t)$ is the (unique) solution to the system:

$$y_j'(t) = f_j(t, y_1(t), \ldots, y_a(t)), \qquad y_j(0) = \tilde{y}_j \tag{11}$$

for all $t \in [0, \sigma]$, where σ is any positive value.

Theorem 3. *Suppose that for each $1 \leq j \leq a$ we have a sequence of random variables $(Z_j(i))_{i=0}^{\infty}$ which is adapted to some filtration $(\mathcal{H}_i)_{i=0}^{\infty}$. Let $n \in \mathbb{N}$, and $\beta, b, \lambda, \delta > 0$ be any parameters such that $\lambda \geq \max\left\{\beta + B, \frac{L+BL+\delta n}{3L}\right\}$. Given an arbitrary stopping time $I \geq 0$ adapted to $(\mathcal{H}_i)_{i=0}^{\infty}$, suppose that the following properties hold for each $0 \leq i < \min\{I, \sigma n\}$:*

1. *The 'Boundedness Hypothesis': $\max_j |\Delta Z_j(i)| \leq \beta$, and $\max_j \mathbb{E}[(\Delta Z_j(i))^2 \mid \mathcal{H}_i] \leq b$*

2. *The 'Trend Hypothesis': $(i/n, Z_1(i)/n, \ldots Z_a(i)/n) \in \mathcal{D}$ and*

$$\mathbb{E}[\Delta Z_j(i) \mid \mathcal{H}_i] \leq f_j(i/n, Z_1(i)/n, \ldots Z_a(i)/n) + \delta$$

for each $1 \leq j \leq a$.

3. *The 'Initial Condition': $Z_j(0) \leq y_j(0)n + \lambda$ for all $1 \leq j \leq a$.*

Then, with probability at least $1 - 2a\exp\left(-\frac{\lambda^2}{2(b\sigma n + 2\beta\lambda)}\right)$,

$$Z_j(i) \leq ny_j(i/n) + 3\lambda e^{2Li/n} \tag{12}$$

for all $1 \leq j \leq a$ and $0 \leq i \leq \min\{I, \sigma n\}$.

Remark 1 (Simplified Parameters)**.** By taking $b = \beta^2$ and $I = \lceil \sigma n \rceil$, we can recover a simpler version of the theorem which is sufficient for many applications, including the motivating example of Section 1.1.

Remark 2 (Stopping Time Selection)**.** Let $0 \leq \gamma \leq 1$ be an additional parameter to Theorem 3. The stopping time I is most commonly applied in the following way. Suppose that $(\mathcal{E}_i)_{i=0}^{\infty}$ is a sequence of events adapted to $(\mathcal{H}_i)_{i=0}^{\infty}$, and for each $0 \leq i < \sigma n$, Conditions 1. and 2. are *only* verified when $\mathcal{E}_i$ holds. Moreover, assume that $\mathbb{P}[\cap_{i=0}^{\sigma n-1} \mathcal{E}_i] = 1 - \gamma$. By defining I to be the smallest $i \geq 0$ such that $\mathcal{E}_i$ does *not* occur, Theorem 3 implies that with probability at least $1 - 2a \exp\left(-\frac{\lambda^2}{2(b\sigma n + 2\beta\lambda)}\right) - \gamma$,

$$Z_j(i) \leq ny_j(i/n) + 3\lambda e^{2Li/n} \tag{13}$$

for all $1 \leq j \leq a$ and $0 \leq i \leq \sigma n$.

Remark 3 (Competitive Functions)**.** Theorem 3 yields upper bounds for families of random variables. There is a symmetric theorem for lower bounds, where all the appropriate inequalities are reversed and the functions f_j are competitive instead of cooperative.

We conclude the section with a corollary of Theorem 3 which provides a useful extension of the theorem. Roughly speaking, the extension says that when verifying Conditions 1. and 2. at time $0 \leq i \leq \min\{\sigma n, I\}$, it does not hurt to assume that (12) holds.

Corollary 4 (of Theorem 3)**.** *Suppose that in the terminology of Theorem 3, Conditions 1. and 2. are only verified for each $0 \leq i \leq \min\{I, \sigma n\}$ which satisfies $Z_{j'}(i) \leq ny_{j'}(i/n) + 3\lambda e^{2Li/n}$ for all $1 \leq j' \leq a$. In this case, the conclusion of Theorem 3 still holds. I.e., with probability at least $1 - 2a \exp\left(-\frac{\lambda^2}{2(b\sigma n + 2\beta\lambda)}\right)$,*

$$Z_j(i) \leq ny_j(i/n) + 3\lambda e^{2Li/n}$$

for all $1 \leq j \leq a$ and $0 \leq i \leq \min\{I, \sigma n\}$.

Proof of Corollary 4. Let I^* be the first $i \geq 0$ such that

$$Z_{j'}(i) > ny_{j'}(i/n) + 3\lambda e^{2Li/n}$$

for some $1 \leq j' \leq a$. Clearly, I^* is a stopping time adapted to $(\mathcal{H}_i)_{i=0}^{\infty}$. Moreover, by the assumptions of the corollary, Conditions 1. and 2. hold for each $0 \leq i \leq \min\{I^*, I, \sigma n\}$ and $1 \leq j \leq a$. Thus, Theorem 3 implies that with probability at least $1 - 2a \exp\left(-\frac{\lambda^2}{2(b\sigma n + 2\beta\lambda)}\right)$,

$$Z_j(i) \leq ny_j(i/n) + 3\lambda e^{2Li/n}$$

for all $1 \leq j \leq a$ and $0 \leq i \leq \min\{I, I^*, \sigma n\}$. Since the preceding event holds if and only if $I^* > \min\{I, \sigma n\}$, the corollary is proven.

□

3 Proving Theorem 3

Before proceeding to the proof of Theorem 3, we provide some intuition for our approach by presenting a proof of the deterministic setting (i.e., Theorem 2). The notation and structure of the proof is intentionally chosen so as to relate to the analogous approach taken in the proof of Theorem 3. Moreover, the main claim we prove can be viewed as an *approximate* version of Theorem 2, in which the upper bounds on $z_j(0)$ and z_j' only hold up to an additive constant $\delta > 0$.

Proof of Theorem 2. Let us assume that $c = 0$ is the boundary of the domain, and L is a Lipschitz constant for the cooperative functions $(f_j)_{j=1}^a$. We shall prove the following: Given an arbitrary $\delta > 0$, if

$$z_j'(t) \leq f(t, z_j(t)) + \delta, \qquad z_j(0) \leq y_j(0) + \delta \tag{14}$$

for all $1 \leq j \leq a$ and $t \geq 0$, then

$$z_j(t) \leq y_j(t) + \delta e^{Lt} \tag{15}$$

for each $1 \leq j \leq a$ and $t \geq 0$. Since (14) holds *each* $\delta > 0$ under the assumptions of Theorem 2, so must (15). This will imply that $z_j(t) \leq y_j(t)$ for each $1 \leq j \leq a$ and $t \geq 0$, thus completing the proof.

In order to prove that (14) implies (15), define

$$g(t) := 2\delta e^{Lt}, \qquad s_j(t) := z_j(t) - (y_j(t) + g(t)), \qquad I_j(t) := [y_j(t), y_j(t) + g(t)).$$

It suffices to show that $\max_{1\leq j\leq a} s_j(t) \leq 0$ for all $t \geq 0$. Observe first that $s_j(0) = z_j(0) - y_j(0) - g(0) \leq -\delta < 0$ for all $1 \leq j \leq a$. Suppose for the sake of contradiction that there exists some $1 \leq j' \leq a$ such that $s_{j'}(t) > 0$ for some $t > 0$. In this case, there must be some value t_1 with $s_{j'}(t_1) = 0$ and $\max_{1\leq j\leq a} s_j(t) < 0$ for all $t < t_1$. Furthermore, there must be some $t_0 < t_1$ such that $s_{j'}(t) \in [-g(t), 0)$ for all $t_0 \leq t < t_1$. Thus, for $t_0 \leq t < t_1$ we have that

$$-g(t) \leq z_{j'}(t) - [y_{j'}(t) + g(t)] < 0 \qquad \Longrightarrow \qquad y_{j'}(t) \leq z_{j'}(t) < y_{j'}(t) + g(t), \tag{16}$$

and so

$$\begin{aligned} f_{j'}\Big(t, z_1(t), \ldots z_a(t)\Big) &\leq f_{j'}\Big(t, y_1(t) + g(t), \ldots, z_{j'}(t), \ldots, y_a(t) + g(t)\Big) \\ &\leq f_{j'}\Big(t, y_1(t), \ldots, y_a(t)\Big) + Lg(t) \end{aligned} \tag{17}$$

where the first line is by cooperativity of the functions f_j and the second line is by the Lipschitzness of $f_{j'}$ applied to (16). As such, for all $t \in [t_0, t_1)$,

$$\begin{aligned} s_{j'}'(t) &= z_{j'}'(t) - y_{j'}'(t) - g'(t) \\ &= f_{j'}\Big(t, z_1(t), \ldots, z_a(t)\Big) - f_{j'}\Big(t, y_1(t), \ldots, y_a(t)\Big) - g'(t) \\ &\leq Lg(t) - g'(t) = 0 \end{aligned}$$

where the last line uses (17). But now we have a contradiction: $s_{j'}(t_0) \in [-g(t_0), 0)$ so it is negative, $s_{j'}'(t) \leq 0$ on $[t_0, t_1)$, and yet $s_{j'}(t_1) = 0$.

□

Our proof of Theorem 3 is based partly on the *critical interval method.* Similar ideas were used by used by Telcs, Wormald and Zhou [10] as well as Bohman, Frieze and Lubetzky [2] (whose terminology we use here). For a gentle discussion of the method see the paper of the first author and Dudek [1]. Roughly speaking, the critical interval method allows us to assume we have good estimates of key variables during the very steps that we are most concerned with those variables. Historically this method has been used with more standard applications of the differential equation method in order to exploit *self-correcting* random variables, i.e. a variable with the property that when it strays significantly from its trajectory, its expected one-step change makes it likely to move back toward its trajectory. For such a random variable, knowing that it lies in an interval strictly above (or below) the trajectory gives us a more favorable estimate for its expected one-step change. In our setting we use the method for a similar but different reason. In particular since we can only hope for one-sided bounds, we may as well ignore our random variables when they are far away from their bounds (in any case, we do not have or need good estimates for their expected one-step changes etc. during the steps when all variables are far from their bounds).

We give an analogy. A rough proof sketch for Theorem 2 is as follows: in order to have $z_j(t) > y_j(t)$ for some t there must be some time interval during which $z_j \approx y_j$ and during that interval z_j must increase significantly faster than y_j, which contradicts what we know about their derivatives. An analogous proof sketch for Theorem 3 is as follows: in order for $Z_j(i)$ to violate its upper bound, it must first enter a critical interval which we will define to be near the upper bound, and then Z_j must increase significantly (more than we expect it to) over the subsequent steps, which while possible, is unlikely.

Proof of Theorem 3. Fix $0 \le i \le \sigma n$, and set $m := \sigma n$, $t = t_i = i/n$, and $g(t) := 3\lambda e^{2Lt}$ for convenience. Define

$$S_j(i) : = Z_j(i) - (ny_j(t) + g(t)), \quad X_j(i) := \sum_{k=0}^{i-1} \mathbb{E}[\Delta S_j(k) \mid \mathcal{H}_k],$$

$$M_j(i) : = S_j(0) + \sum_{k=0}^{i-1} \left(\Delta S_j(k) - \mathbb{E}[\Delta S_j(k) \mid \mathcal{H}_k]\right),$$

so that $S_j(i) = X_j(i) + M_j(i)$, $(M_j(i))_{i=0}^m$ is a martingale and $X_j(i)$ is $\mathcal{H}_{i-1}$-measurable (i.e. $(X_j(i) + M_j(i))_{i=0}^m$ is the Doob decomposition of $(S_j(i))_{i=0}^m$). Note that we can view $S_j(i)/n$ as the random analogue of $s_j(t) = m_j(t) + x_j(t)$ from the proof of Theorem 2. In the previous deterministic setting, the decomposition $s_j(t) = m_j(t) + x_j(t)$ is redundant, as $m_j(t) = s_j(0)$, and so $x_j(t)$ and $s_j(t)$ differ by a constant. In contrast, $M_j(i)$ is $S_j(0)$, *plus* $\sum_{k=0}^{i-1} (\Delta S_j(k) - \mathbb{E}[\Delta S_j(k) \mid \mathcal{H}_k])$, the latter of which we can view as introducing some *random noise.* We handle this random noise by showing that $M_j(i)$ is typically concentrated about $S_j(0)$ due to Freedman's inequality (see Theorem 7 in Appendix A). We refer to this as the *probabilistic* part of the proof. Assuming that this concentration holds, we can upper bound $Z_j(i)$ by $ny_j(t) + g(t)$ via an argument which proceeds analogously to the proof of Theorem 2. This is the *deterministic* part of the proof.

Beginning with the probabilistic part of the proof, we restrict our attention to $0 \le i <$

$\min\{I, m\}$. Observe first that

$$\begin{aligned}\Delta M_j(i) &= \Delta S_j(i) - \mathbb{E}[\Delta S_j(i) \mid \mathcal{H}_i] \\ &= \Delta[Z_j(i) - (ny_j(t) + g(t))] - \mathbb{E}[\Delta[Z_j(i) - (ny_j(t) + g(t))] \mid \mathcal{H}_i] \\ &= \Delta Z_j(i) - \mathbb{E}[\Delta Z_j(i) \mid \mathcal{H}_i], \end{aligned} \tag{18}$$

and so by Condition 1. ,

$$|\Delta M_j(i)| \leq |\Delta Z_j(i)| + |\mathbb{E}[\Delta Z_j(i) \mid \mathcal{H}_i]| \leq 2\beta.$$

Also, $\mathbf{Var}[\Delta M_j(i) \mid \mathcal{H}_i] = \mathbb{E}[(\Delta Z_j(i) - \mathbb{E}[\Delta Z_j(i) \mid \mathcal{H}_i])^2 \mid \mathcal{H}_i]$. Thus,

$$\begin{aligned}\mathbf{Var}[\Delta M_j(i) \mid \mathcal{H}_i] &= \mathbb{E}[\Delta Z_j(i)^2 \mid \mathcal{H}_i] - \mathbb{E}[\Delta Z_j(i) \mid \mathcal{H}_i]^2 \\ &\leq \mathbb{E}[\Delta Z_j(i)^2 \mid \mathcal{H}_i] \\ &\leq b \qquad\qquad \text{by Condition 1.}\end{aligned}$$

We can therefore apply Theorem 7 to get that

$$\mathbb{P}(\exists\, 0 \leq j \leq a, 0 \leq i \leq \min\{I, m\} : |M_j(i) - M_j(0)| \geq \lambda) \leq 2a \exp\left(-\frac{\lambda^2}{2(bm + 2\beta\lambda)}\right). \tag{19}$$

Suppose the above event does *not* happen i.e., for all $0 \leq j \leq a, 0 \leq i \leq \min\{m, I\}$ we have that $|M_j(i) - M_j(0)| < \lambda$. We will show that we also have $Z_j(i) \leq ny_j(t) + g(t)$ for all $0 \leq i \leq \min\{m, I\}$ and $1 \leq j \leq a$ (equivalently, $\max_j S_j(i) \leq 0$ for all $0 \leq i \leq \min\{m, I\}$). This implication is the deterministic part of the proof. By combining it with the probability bound of (19), this will complete the proof of Theorem 3.

Suppose for the sake of contradiction that i' is the minimal integer such that $0 \leq i' \leq \min\{m, I\}$ and $Z_j(i') > ny_j(t_{i'}) + g(t_{i'})$ for some j. Define the **critical interval**

$$I_j(i) := [ny_j(t), ny_j(t) + g(t)]\,.$$

First observe that since $g(0) := 3\lambda > \lambda$, Condition 3. implies that $i' > 0$ (and so $i' - 1 \geq 0$.) We claim that $Z_j(i'-1) \in I_j(i'-1)$. Indeed, note that by the minimality of i' we have that $Z_j(i'-1) \leq ny_j(t_{i'-1}) + g(t_{i'-1})$. On the other hand, $|y_j'| = |f_j| \leq B$ and so each y_j is B-Lipschitz. Thus, since $\lambda \geq \beta + B$ (by assumption),

$$\begin{aligned}Z_j(i'-1) \geq Z_j(i') - \beta &> ny_j(t_{i'}) + g(t_{i'}) - \beta \\ &\geq ny_j(t_{i'-1}) + 3\lambda - \beta - B \\ &\geq ny_j(t_{i'-1}).\end{aligned}$$

As a result, $Z_j(i'-1) \in I_j(i'-1)$. Now let $i'' \leq i'-1$ be the minimal integer with the property that for all $i'' \leq i \leq i'-1$, we have that $Z_j(i) \in I_j(i)$. Then $Z_j(i''-1) \notin I_j(i''-1)$ and by the minimality of i' we must have that $Z_j(i''-1) < ny_j(t_{i''-1})$. By once again using the fact that y_j is B-Lipschitz,

$$Z_j(i'') \leq Z_j(i''-1) + \beta < ny_j(t_{i''-1}) + \beta \leq ny_j(t_{i''}) + \beta + B. \tag{20}$$

Now, since $Z_j(i') > ny_j(t_{i'}) + g(t_{i'})$, we can apply (20) to get that

$$\begin{aligned} S_j(i') - S_j(i'') &= (Z_j(i') - ny_j(t_{i'}) - g(t_{i'})) - (Z_j(i'') - ny_j(t_{i''}) - g(t_{i''})) \\ &> g(t_{i''}) - \beta - B \\ &\geq 3\lambda - \beta - B. \end{aligned} \tag{21}$$

Intuitively, (21) says that $S_j(i)$ increases significantly between steps i'' and i'. We will now argue that $\mathbb{E}[\Delta S_j(i)]$ is nonpositive between steps i'' and i'. Informally, by scaling by n and interpreting $\mathbb{E}[\Delta S_j(i)]$ as the "derivative" of $S_j(i)/n$ evaluated at i/n, this will allow us to derive a contradiction in an analogous way as in the final part of the proof of Theorem 2.

Observe first that for each $i'' \leq i \leq i' - 1$, we have that

$$\begin{aligned} \mathbb{E}[\Delta S_j(i) \mid \mathcal{H}_i] &= \mathbb{E}[\Delta Z_j(i) \mid \mathcal{H}_i] - \Delta n y_j(t) - \Delta g(t) \\ &\leq f_j(t, Z_1(i)/n, \ldots Z_a(i)/n) + \delta \\ &\qquad - f_j(t, y_1(t), \ldots, y_a(t)) + (L + BL)n^{-1} - n^{-1}g'(t) \end{aligned} \tag{22}$$

where the first line is by definition and line (22) will now be justified. The first two terms follow since by Condition 2, $(t, Z_1(i)/n, \ldots Z_a(i)/n) \in \mathcal{D}$, and

$$\mathbb{E}[\Delta Z_j(i) \mid \mathcal{H}_i] \leq f_j(t, Z_1(i)/n, \ldots Z_a(i)/n) + \delta.$$

For the third and fourth terms of (22), note that

$$\begin{aligned} \Delta n y_j(t) = n[y_j(t_{i+1}) - y_j(t_i)] &= n \int_{t_i}^{t_{i+1}} y_j'(\tau)\, d\tau \\ &= n \int_{t_i}^{t_{i+1}} f_j(\tau, y_1(\tau), \ldots, y_a(\tau))\, d\tau \\ &\geq n \int_{t_i}^{t_{i+1}} f_j(t, y_1(t_i), \ldots, y_a(t_i)) - L|\tau - t_i| - L|y_j(\tau) - y_j(t_i)|\, d\tau \\ &\geq n \int_{t_i}^{t_{i+1}} f_j(t, y_1(t_i), \ldots, y_a(t_i)) - (L + BL)|t - t_i|\, dt \\ &\geq f_j(t, y_1(t_i), \ldots, y_a(t_i)) - (L + BL)n^{-1} \end{aligned}$$

For the last term of (22), we have that

$$\begin{aligned} \Delta g(t) = 3\lambda \left(e^{2Lt_{i+1}} - e^{2Lt_i}\right) &= 3\lambda e^{2Lt_i} \left(e^{2L/n} - 1\right) \\ &\geq 3\lambda e^{2Lt_i} \left(\frac{2L}{n}\right) = n^{-1} g'(t). \end{aligned}$$

Observe now that by cooperativity, $f_j(t, Z_1(i)/n, \ldots Z_a(i)/n)$ is upper bounded by

$$f_j\left(t, \frac{ny_1(t) + g(t)}{n}, \ldots, \frac{ny_{j-1}(t) + g(t)}{n}, \frac{Z_j(i)}{n}, \frac{ny_{j+1}(t) + g(t)}{n}, \ldots, \frac{ny_a(t) + g(t)}{n}\right). \tag{23}$$

Now, since $Z_j(i) \in I_j(i)$, we can apply the Lipschitzness of f_j to (23) to get that

$$f_j(t, Z_1(i)/n, \ldots Z_a(i)/n) \leq f_j(t, y_1(t), \ldots, y_a(t)) + Lg(t)/n.$$

As such, applied to (22),

$$\begin{aligned}
\mathbb{E}[\Delta S_j(i) \mid \mathcal{H}_i] &\leq f_j(t, Z_1(i)/n, \ldots Z_a(i)/n) + \delta - f_j(t, y_1(t), \ldots, y_a(t)) + (L+BL)n^{-1} - n^{-1}g'(t) \\
&\leq Ln^{-1}g(t) + \delta - n^{-1}g'(t) + (L+BL)n^{-1} \\
&= Ln^{-1}g(t) + \delta - n^{-1}2Lg(t) + (L+BL)n^{-1} \\
&\leq -[Lg(t) - (L+BL+\delta n)]n^{-1} \\
&\leq -[3L\lambda - (L+BL+\delta n)]n^{-1} \qquad (24) \\
&\leq 0, \qquad (25)
\end{aligned}$$

where the final line follows since $\lambda \geq \frac{L+BL+\delta n}{3L}$.

Therefore, for $i'' \leq i \leq i'-1$ we have that

$$0 \geq \mathbb{E}[\Delta S_j(i) \mid \mathcal{H}_i] = \mathbb{E}[\Delta X_j(i) \mid \mathcal{H}_i] + \mathbb{E}[\Delta M_j(i) \mid \mathcal{H}_i] = \Delta X_j(i)$$

since $(M_j(i))_{i=0}^m$ is a martingale and $\Delta X_j(i)$ is $\mathcal{H}_i$-measurable. In particular,

$$X_j(i') \leq X_j(i''). \qquad (26)$$

At this point, we use the event we assumed regarding $(M_j(i))_{i=0}^m$ (directly following (19)) to get that

$$M_j(i') - M_j(i'') \leq |M_j(i') - M_j(0)| + |M_j(i'') - M_j(0)| \leq 2\lambda. \qquad (27)$$

But now we can derive our final contradiction (explanation follows):

$$\begin{aligned}
3\lambda - \beta - B &< S_j(i') - S_j(i'') \\
&= X_j(i') - X_j(i'') + M_j(i') - M_j(i'') \\
&\leq 2\lambda.
\end{aligned}$$

Indeed the first line is from (21), the second is by the Doob decomposition, and the last line follows from (26) and (27). But the last line is a contradiction since we chose $\lambda \geq \beta + B$.

□

4 Weakening the Assumptions of Theorem 2

There are additional assumpions we can make when we check Conditions a)-c) on the f_j and Conditions 1. and 2. of Theorem 2. We will list these assumptions below. The fact that it suffices to check the conditions under these assumptions follows from checking that our proof only uses the conditions when the assumptions hold.

- Condition a): f_j only needs to be L-Lipschitz on the set of points

  $$\mathcal{D}^* := \Big\{(t, z_1, \ldots, z_a) \in \mathbb{R}^{a+1} : 0 \leq t \leq \sigma, y_{j'}(t) \leq z_{j'} \leq y_{j'}(t) + g(t) \text{ for } 1 \leq j' \leq a\Big\} \subseteq \mathcal{D}.$$

- Condition b): We only use this condition to conclude that the system (11) has a unique solution and that those solutions y_j are B-Lipschitz. So, it suffices to just check directly that (11) has a unique solution that is B-Lipschitz.

- Condition c): It suffices to have that $f_j(t, z_1, \dots z_a)$ is upper bounded by

$$f_j\left(t, \frac{ny_1(t)+g(t)}{n}, \dots, \frac{ny_{j-1}(t)+g(t)}{n}, z_j, \frac{ny_{j+1}(t)+g(t)}{n}, \dots, \frac{ny_a(t)+g(t)}{n}\right)$$

whenever $z_{j'} \leq y_{j'}(t) + g(t)/n$ for all j' and $z_j \leq y_j(t)$.

- Condition 2. : It suffices to have the following for each $1 \leq j \leq a$. If $Z_{j'}(i) \leq ny_{j'}(i/n) + g(i/n)$ for all $1 \leq j' \leq a$ and $Z_j \geq ny_j(i/n)$, then $(i/n, Z_1(i)/n, \dots Z_a(i)/n) \in \mathcal{D}$ and

$$\mathbb{E}[\Delta Z_j(i) \mid \mathcal{H}_i] \leq f_j(i/n, Z_1(i)/n, \dots Z_a(i)/n) + \delta$$

5 Recovering a Version of Wormald's Theorem

In this section we recover the standard (two-sided) differential equation method of Wormald [13]. The statement resembles the recent version given by Warnke [12] in the sense that it does not use any asymptotic notation and instead gives explicit bounds for error estimates and failure probabilities. Like Warnke's proof, ours has a *probabilistic part* and a *deterministic part*. Our probabilistic part is much the same as Warnke's in that we apply a deviation inequality (though we use Freedman's theorem rather than the Azuma-Hoeffding inequality) to the martingale part of a Doob decomposition. That being said, the deterministic part of our argument is quite different than the deterministic part of Warnke's argument. In fact, we were not able to adapt Warnke's argument to the one-sided setting.

Given $a \in \mathbb{N}$, suppose that $\mathcal{D} \subseteq \mathbb{R}^{a+1}$ is a bounded domain, and for $1 \leq j \leq a$, let $f_j : \mathcal{D} \to \mathbb{R}$. We assume that the following hold for each j:

1. f_j is L-Lipschitz, and

2. $|f_j| \leq B$ on $\mathcal{D}$.

Given $(0, \tilde{y}_1, \dots, \tilde{y}_a) \in \mathcal{D}$, assume that $y_1(t), \dots, y_a(t)$ is the (unique) solution to the system

$$y_j'(t) = f_j(t, y_1(t), \dots, y_a(t)), \qquad y_j(0) = \tilde{y}_j. \tag{28}$$

for all $t \in [0, \sigma]$ where σ is a positive value. Note that unlike in Theorem 3, we make a further assumption involving σ below in Theorem 5.

Theorem 5. *Suppose for each $1 \leq j \leq a$ we have a sequence of random variables $(Y_j(i))_{i=0}^{\infty}$ which is adapted to the filtration $(\mathcal{H}_i)_{i=0}^{\infty}$. Let $n \in \mathbb{N}$, and $\beta, b, \lambda, \delta > 0$ be any parameters such that $\lambda \geq \frac{L+BL+\delta n}{3L}$. Moreover, assume that $\sigma > 0$ is any value such that $(t, y_1(t), \dots, y_a(t))$ has ℓ^∞-distance at least $3\lambda e^{2Lt}/n$ from the boundary of $\mathcal{D}$ for all $t \in [0, \sigma)$. Given an arbitrary stopping time $I \geq 0$ adapted to $(\mathcal{H}_i)_{i=0}^{\infty}$, suppose that the following properties hold for each $0 \leq i < \min\{I, \sigma n\}$:*

1. *The 'Boundedness Hypothesis': $\max_j |\Delta Y_j(i)| \leq \beta$, and $\max_j \mathbb{E}[(\Delta Y_j(i))^2 \mid \mathcal{H}_i] \leq b$.*

2. *The 'Trend Hypothesis': If* $(i/n, Y_1(i)/n, \dots Y_a(i)/n) \in \mathcal{D}$*, then*

$$\Big|\mathbb{E}[\Delta Y_j(i) \mid \mathcal{H}_i] - f_j(i/n, Y_1(i)/n, \dots Y_a(i)/n)\Big| \le \delta$$

   for each $1 \le j \le a$*.*

3. *The 'Initial Condition':* $|Y_j(0) - y_j(0)n| \le \lambda$ *for all* $1 \le j \le a$*.*

Then, with probability at least $1 - 2a\exp\left(-\frac{\lambda^2}{2(b\sigma n + 2\beta\lambda)}\right)$*,*

$$|Y_j(i) - ny_j(i/n)| \le 3\lambda e^{2Li/n} \tag{29}$$

for all $1 \le j \le a$ *and* $0 \le i \le \min\{I, \sigma n\}$*.*

We conclude the section with an extension of Theorem 5 analogous to Corollary 4 of Theorem 3. We omit the proof, as it follows almost identically to the proof of Corollary 4.

Corollary 6 (of Theorem 5)**.** *Suppose that in the terminology of Theorem 5, Conditions 1. and 2. are only verified for each* $0 \le i \le \min\{I, \sigma n\}$ *which satisfies* $|Y_{j'}(i) - ny_{j'}(i/n)| \le 3\lambda e^{2Li/n}$ *for all* $1 \le j' \le a$*. In this case, the conclusion of Theorem 5 still holds. I.e., with probability at least* $1 - 2a\exp\left(-\frac{\lambda^2}{2(b\sigma n + 2\beta\lambda)}\right)$*,*

$$|Y_j(i) - ny_j(i/n)| \le 3\lambda e^{2Li/n}$$

for all $1 \le j \le a$ *and* $0 \le i \le \min\{I, \sigma n\}$*.*

Proof of Theorem 5. Fix $0 \le i \le \sigma n$, and again set $m := \sigma n$, $t = t_i = i/n$, and $g(t) := 3\lambda e^{2Lt}$ for convenience. Define

$$\begin{aligned}
S_j^{\pm}(i) &:= Y_j(i) - (ny_j(t) \pm g(t)),\\
X_j^{\pm}(i) &:= \sum_{k=0}^{i-1} \mathbb{E}[\Delta S_j^{\pm}(k) \mid \mathcal{H}_k],\\
M_j^{\pm}(i) &:= S_j^{\pm}(0) + \sum_{k=0}^{i-1} \left(\Delta S_j^{\pm}(k) - \mathbb{E}[\Delta S_j^{\pm}(k) \mid \mathcal{H}_k]\right)
\end{aligned}$$

so that $(X_j^{\pm}(i) + M_j^{\pm}(i))_{i=0}^m$ is the Doob decomposition of $(S_j^{\pm}(i))_{i=0}^m$. Note that

$$\Delta S_j^{\pm}(k) - \mathbb{E}[\Delta S_j^{\pm}(k) \mid \mathcal{H}_k] = \Delta Y_j(k) - \mathbb{E}[\Delta Y_j(k) \mid \mathcal{H}_k],$$

and so $M_j^{+}(i)$ is almost the same as $M_j^{-}(i)$. More precisely, we have

$$M_j^{\pm}(i) = M_j(i) \mp g(0)$$

where

$$M_j(i) := Y_j(0) - ny_j(0) + \sum_{k=0}^{i-1} (\Delta Y_j(k) - \mathbb{E}[\Delta Y_j(k) \mid \mathcal{H}_k])$$

(which is also a martingale). As in the proof of Theorem 3, we have $|\Delta M_j(i)| \le 2\beta$ and $\mathbf{Var}[\Delta M_j(i) \mid \mathcal{H}_i] \le b$. Now, by Theorem 7 we have that

$$\mathbb{P}\Big(\exists\, 0 \le j \le a, 0 \le i \le m \text{ such that } |M_j(i) - M_j(0)| \ge \lambda\Big) \le 2a \exp\left(-\frac{\lambda^2}{2(bm + 2\beta\lambda)}\right). \tag{30}$$

Suppose that the event above does not happen, so for all $0 \le j \le a, 0 \le i \le m$ we have that $|M_j(i) - M_j(0)| < \lambda$. We will show that we also have $|Y_j(i) - ny_j(t)| \le g(t)$ for all $0 \le i \le m$. Note that y_j is B-Lipschitz as before. Define the **critical interval**

$$I_j(i) := [ny_j(t) - g(t), ny_j(t) + g(t)]\,.$$

Suppose for the sake of contradiction that i' is minimal with $0 \le i' \le m$ and $Y_j(i') \notin I_j(i')$ for some j. We will consider the case where $Y_j(i') > ny_j(t) + g(t)$ (the case where $Y_j(i') < ny_j(t) - g(t)$ is handled similarly with some inequalities reversed). In other words, $S_j^+(i') > 0$. First observe that since $g(0) := 3\lambda$, Condition 3. implies $S_j^+(0) \le -2\lambda$. In particular, $i' > 0$ and

$$S_j^+(i') - S_j^+(0) > 2\lambda. \tag{31}$$

For $0 \le i < i'$, we have (explanation follows)

$$\begin{aligned}
\mathbb{E}[\Delta S_j^+(i) \mid \mathcal{H}_i] &= \mathbb{E}[\Delta Y_j(i) \mid \mathcal{H}_i] - \Delta n y_j(t) - \Delta g(t)\\
&\le f_j(t, Y_1(i)/n, \dots Y_a(i)/n) + \delta - f_j(t, y_1(t), \dots, y_a(t)) + (L + BL)n^{-1} - n^{-1}g'(t)\\
&\le Ln^{-1}g(t) + \delta + (L + BL)n^{-1} - n^{-1}g'(t)\\
&\le -[3L\lambda - (L + BL + \delta n)]n^{-1}\\
&\le 0.
\end{aligned}$$

Indeed, the first line is by definition and the second line follows just like (22), with the minor caveat that we now must show that $(t, Y_1(i)/n, \dots, Y_a(i)/n)$ is within the domain $\mathcal{D}$ to apply Condition 2 (recall that in Theorem 3 this is simply assumed). Observe that by the definition of σ, $(t, y_1(t), \dots, y_a(t))$ is in $\mathcal{D}$ *and* at least ℓ^∞-distance $g(t)/n$ from the boundary of $\mathcal{D}$. On the other hand, since $Y_{j'}(i) \in I_{j'}(i)$ for all $1 \le j' \le a$, we know that $|Y_{j'}(i)/n - y_{j'}(t)| \le g(t)/n$ for all $1 \le j' \le a$. Thus, $(t, Y_1(i)/n, \dots Y_a(i)/n) \in \mathcal{D}$, and so

$$\mathbb{E}[\Delta Y_j(i) \mid \mathcal{H}_i] \le f_j(t, Y_1(i)/n, \dots, Y_a(i)/n) + \delta$$

by Condition 2. The remaining justification is much the same as before. Since f_j is L-Lipschitz and $|Y_{j'}(i) - ny_{j'}(t)| \le g(t)$ for all j', we have $f_j(t, Y_1(i)/n, \dots Y_a(i)/n) \le f_j(t, y_1(t), \dots, y_a(t)) + Ln^{-1}g(t)$ and the third line follows. The fourth and fifth lines follow just as (24) and (25). Therefore, for $0 \le i < i'$ we have that

$$0 \ge \mathbb{E}[\Delta S_j^+(i) \mid \mathcal{H}_i] = \mathbb{E}[\Delta X_j^+(i) \mid \mathcal{H}_i] + \mathbb{E}[\Delta M_j^+(i) \mid \mathcal{H}_i] = \Delta X_j^+(i)$$

since $(M_j^+(i))_{i=0}^\infty$ is a martingale and $\Delta X_j^+(i)$ is $\mathcal{H}_i$-measurable. In particular

$$X_j^+(i') \le X_j^+(0). \tag{32}$$

But now we can derive our final contradiction (explanation follows):

$$\begin{aligned}2\lambda &< S_j^+(i') - S_j^+(0) \\ &= X_j^+(i') - X_j^+(0) + M_j(i') - M_j(0) \\ &\leq \lambda.\end{aligned}$$

Indeed, the first line is from (31), the second line is by the Doob decomposition, and the last follows from (32) and our assumption that the event described on line (30) does not happen. Of course $2\lambda < \lambda$ is a contradiction and we are done. □

References

[1] Patrick Bennett and Andrzej Dudek. A gentle introduction to the differential equation method and dynamic concentration. *Discrete Math.*, 345(12):Paper No. 113071, 17, 2022.

[2] Tom Bohman, Alan Frieze, and Eyal Lubetzky. Random triangle removal. *Adv. Math.*, 280:379–438, 2015.

[3] Elaheh Fata, Will Ma, and David Simchi-Levi. Multi-stage and multi-customer assortment optimization with inventory constraints. *CoRR*, abs/1908.09808, 2019.

[4] Moran Feldman, Ola Svensson, and Rico Zenklusen. Online contention resolution schemes with applications to bayesian selection problems. *SIAM Journal on Computing*, 50(2):255–300, 2021.

[5] David A. Freedman. On Tail Probabilities for Martingales. *The Annals of Probability*, 3(1):100 – 118, 1975.

[6] Hu Fu, Zhihao Gavin Tang, Hongxun Wu, Jinzhao Wu, and Qianfan Zhang. Random Order Vertex Arrival Contention Resolution Schemes for Matching, with Applications. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 68:1–68:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

[7] Euiwoong Lee and Sahil Singla. Optimal Online Contention Resolution Schemes via Ex-Ante Prophet Inequalities. In Yossi Azar, Hannah Bast, and Grzegorz Herman, editors, *26th Annual European Symposium on Algorithms (ESA 2018)*, volume 112 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 57:1–57:14, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[8] Calum MacRury, Will Ma, and Nathaniel Grammel. On (random-order) online contention resolution schemes for the matching polytope of (bipartite) graphs. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 1995–2014. SIAM, 2023.

[9] M. Michel Petrovitch. Sur une manière d'étendre le théorème de la moyenne aux équations différentielles du premier ordre. *Mathematische Annalen*, 54:417–436, 1901.

[10] András Telcs, Nicholas Wormald, and Sanming Zhou. Hamiltonicity of random graphs produced by 2-processes. *Random Structures Algorithms*, 31(4):450–481, 2007.

[11] Wolfgang Walter. Differential inequalities and maximum principles: theory, new methods and applications. In *Proceedings of the Second World Congress of Nonlinear Analysts, Part 8 (Athens, 1996)*, volume 30, pages 4695–4711, 1997.

[12] Lutz Warnke. On wormald's differential equation method. *Accepted to Combinatorics, Probability, and Computing*, abs/1905.08928, 2019.

[13] Nicholas C. Wormald. Differential Equations for Random Processes and Random Graphs. *The Annals of Applied Probability*, 5(4):1217 – 1235, 1995.

A Martingale Concentration

Suppose that $(\mathcal{H}_i)_{i=0}^{\infty}$ is an increasing sequence of σ-algebras (i.e., $\mathcal{H}_{i-1} \subseteq \mathcal{H}_i$ for all $i \geq 1$.) Moreover, let $(M_i)_{i=0}^{\infty}$ be a sequence of random variables adapted to $(\mathcal{H}_i)_{i=0}^{\infty}$ (i.e., each M_i is $\mathcal{H}_i$-measurable), and let I be an arbitrary stopping time adapted to $(\mathcal{H}_i)_{i=0}^{\infty}$) (i.e., $\{I = i\}$ is $\mathcal{H}_i$-measurable for each $i \geq 0$). Recall that $\Delta M_i := M_{i+1} - M_i$ and $\mathbf{Var}(\Delta M_i \mid \mathcal{H}_i) := \mathbb{E}[\Delta M_i^2 \mid \mathcal{H}_i] - \mathbb{E}[\Delta M_i \mid \mathcal{H}_i]^2$ for $i \geq 0$.

Theorem 7 (Freedman's Inequality [5])**.** *Fix $m \in \mathbb{N}$ and $\beta, b \geq 0$. Assume that for each $0 \leq i < m$, $\mathbb{E}[\Delta M_i \mid \mathcal{H}_i] = 0$, $|\Delta M_i| \leq \beta$, and $\mathbf{Var}(\Delta M_i \mid \mathcal{H}_i) \leq b$. Then, for any $0 < \varepsilon < 1$,*

$$\mathbb{P}(\exists\, 0 \leq i \leq m : |M_i - M_0| \geq \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{2(bm + \beta\varepsilon)}\right).$$

Moreover, if I is an arbitrary stopping time adapted to $(\mathcal{H}_i)_{i=0}^{\infty}$, and the above conditions regarding $(M_i)_{i=0}^{\infty}$ are only verified for all $0 \leq i < \min\{I, m\}$, then

$$\mathbb{P}(\exists\, 0 \leq i \leq \min\{I, m\} : |M_i - M_0| \geq \varepsilon) \leq 2 \exp\left(-\frac{\varepsilon^2}{2(bm + \beta\varepsilon)}\right).$$