Journal of Mathematical Biology (2021) 82:2
https://doi.org/10.1007/s00285-021-01565-7

Mathematical Biology

Maximizing the total population with logistic growth in a patchy environment

Kentaro Nagahara[1] · Yuan Lou[2] · Eiji Yanagida[1]

Received: 23 February 2020 / Revised: 22 September 2020 / Accepted: 29 December 2020 /
Published online: 19 January 2021

Abstract

This paper is concerned with a nonlinear optimization problem that naturally arises in population biology. We consider the population of a single species with logistic growth residing in a patchy environment and study the effects of dispersal and spatial heterogeneity of patches on the total population at equilibrium. Our objective is to maximize the total population by redistributing the resources among the patches under the constraint that the total amount of resources is limited. It is shown that the global maximizer can be characterized for any number of patches when the diffusion rate is either sufficiently small or large. To show this, we compute the first variation of the total population with respect to resources in the two patches case. In the case of three or more patches, we compute the asymptotic expansion of all patches by using the Taylor expansion with respect to the diffusion rate. To characterize the shape of the global maximizer, we use a recurrence relation to determine all coefficients of all patches.

Keywords Dispersal · Total population · Patchy environment · Difference equation

Mathematics Subject Classification 34H05 · 39A12 · 49J15 · 92D25

YL was partially supported by the NSF grant DMS-1853561; EY was partially supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (A) (No. 17H01095).

✉ Kentaro Nagahara
nagahara@hst.titech.ac.jp

Yuan Lou
lou@math.ohio-state.edu

Eiji Yanagida
yanagida@math.titech.ac.jp

[1] Department of Mathematics, Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8551, Japan

[2] Department of Mathematics, Ohio State University, Columbus, OH 43210, USA

1 Introduction

Understanding the effect of dispersal in heterogeneous environment on population dynamics is an important issue in spatial ecology (Cantrell and Cosner 2003). Generally large diffusion tends to reduce the spatial variations in population distributions, while small diffusion might help organisms adapt to the local environment. In this paper we are interested in the impact of dispersal upon the total population of a single species residing in a spatially heterogeneous patchy environment. More specifically, we ask the following question: Given the total amount of resources, how should we distribute the resources across the habitat in order to maximize the total population of a species?

To address this question, we consider the following system for a single species with logistic growth in a patchy environment:

$$\begin{cases} \dfrac{d}{dt} v_i(t) = v_i(m_i - v_i) + \delta(v_{i-1} + v_{i+1} - 2v_i), & i \in \Omega, t \in \mathbb{R}_+, \\ v_0(t) = v_1(t),\ v_{N+1}(t) = v_N(t), & t \in \mathbb{R}_+, \\ v_i(0) \geq 0,\ \sum_{i=1}^N v_i(0) > 0, & i \in \Omega, \end{cases} \tag{1.1}$$

where $N \geq 2$, $\Omega := \{1, 2, \ldots, N\}$, and $\{m_i\}_{i \in \Omega} \subset \mathbb{R}$ is a sequence which satisfies

$$m_i \geq 0, \quad \sum_{i=1}^N m_i = m > 0. \tag{1.2}$$

The problem (1.1) was first studied by Levins (1969), as a multi-patch model for a single species, where N is the total number of patches and $\delta > 0$ is the diffusion rate. The unknown function $v_i(t)$, $i \in \Omega$, $t \in \mathbb{R}_{\geq 0} := [0, \infty)$, denotes the number of individuals in i-th patch at time t. The constant m_i, $i \in \Omega$, represents the intrinsic growth rate of the species in i-th patch. If $m_i > 0$, then i-th patch is favorable to the species. The second equation in (1.1) means that no individuals cross the boundary of the habitat, so system (1.1) is closed. The constraint (1.2) means that the total amount of resources is limited.

Under assumption (1.2) it is well known that (1.1) has a unique positive steady state, denoted as $\{u_i\}_{i \in \Omega}$, which satisfies

$$\begin{cases} u_i(m_i - u_i) + \delta(u_{i-1} + u_{i+1} - 2u_i) = 0, & i \in \Omega, \\ u_0 = u_1,\ u_{N+1} = u_N. \end{cases} \tag{1.3}$$

Furthermore, as shown in Sect. 3, this unique positive steady state is globally stable and the total population of (1.1) satisfies

$$\sum_{i=1}^N v_i(t) \to \sum_{i=1}^N u_i \quad \text{as } t \to \infty.$$

Our purpose is to maximize the total population $U := \sum_{i=1}^{N} u_i$ at equilibrium under the constraint (1.2). See (Hastings 1982; Holt 1985; Levin 1974; Takeuchi 1989) for related works.

This sort of multi-patch model is called "island chain" model or "stepping stone" model. Such model views the space as a collection of discrete patches. We treat each patch as a point, and view the overall population of a single species as a vector, with each component corresponding to the number of individuals in each patch. Furthermore, we can treat the dispersal in this model as a discrete analogue of the continuous diffusion. For more details, see Allen (1987), Cantrell and Cosner (2003), Hirsch (1984) and references therein. For this reason, this work is closely relevant to the investigation of the following reaction-diffusion equation introduced by Skellam (1951):

$$\begin{cases} v_t = \delta \Delta v + m(x)v - v^2, & (x,t) \in \Omega \times \mathbb{R}_+, \\ \dfrac{\partial v}{\partial \nu} = 0, & (x,t) \in \partial\Omega \times \mathbb{R}_+, \\ v(x,0) \geq 0, \;\; v(x,0) \not\equiv 0, & x \in \overline{\Omega}, \end{cases} \tag{1.4}$$

where $\Omega \subset \mathbb{R}$ is a smooth bounded domain. We also refer to Bai et al. (2016), Cantrell and Cosner (2003), Lou (2006), Lou and Yanagida (2006) and references therein for previous works of (1.4).

The maximization of the total population for the steady state of (1.4) has recently been studied by Mazari et al. (2020) and Nagahara and Yanagida (2018) in $\Omega \subset \mathbb{R}^N$. They showed under some conditions that any global maximizer of the total population for the steady state must be of "bang-bang" type, which gives a partial answer to the conjecture raised by Ding et al. (2010). More recently, Mazari et al. (2020) proved that if $\delta > 0$ is sufficiently large, then the global maximizer is given by $m(x) := \chi_E$, where either $E = (0, m)$ or $(1 - m, 1)$. Their analytical results (Theorem 4, Mazari et al. (2020)) and numerical simulation results indicate that if the diffusion constant is sufficiently small, then fragmentation may occur in the one-dimensional case. However, it is extremely difficult to explicitly determine the maximizer for the steady state of (1.4) in general. We also refer to Mazari et al. (2020) and Mazari and Ruiz-Balet (2020) for related works on PDE models.

This motivates us to study the maximization problem for the difference equation (1.3), for which the computations of the total population can be done (but still fairly non-trivial) for small and large diffusion rates. Our results show that the global maximizer depends crucially on the diffusion rate δ, and the answers are completely different for small δ and large δ. In several cases we are able to show that the global maximizer is of the "bang-bang" type and to determine the maximizers explicitly by finding the specific guiding rules of fragmentation in the multi-patch model (1.3). In particular, fragmentation occurs when the diffusion rate is sufficiently small, which echoes the analytical and numerical findings in Mazari et al. (2020).

In this paper, we do not assume the upper bound for the resource distribution in each patch. There is some difference between the spatially discrete model (1.1) and the continuous model (1.4) in order to have "bang-bang" type of maximizers. For patch model, it follows from (1.2) that $m_i \leq m$ for each i, i.e. an upper bound on the resource distribution in each component is a consequence of the upper bound on the total resource. In contrast, for PDE model (1.4) it is not sufficient to assume that $m \geq 0$ and $\int_\Omega m(x)\,dx$ is bounded by a fixed positive constant, as there must be an upper bound on the resource distribution $m(x)$ to avoid Dirac mass concentrations.

Throughout this paper we will adopt the convention of using a letter to denote a scalar and its boldface to denote a vector or a set, which will be clear from the context.

2 Main results

We define the set

$$\mathcal{M} := \{\{m_i\}_{i\in\Omega} \mid \{m_i\}_{i\in\Omega} \text{ satisfies } (1.2)\} \subset \mathbb{R}^N.$$

For convenience, we express $\{m_i\}_{i\in\Omega}$ by m or $(m_1, m_2, \ldots, m_N)$.

Note that the solution of (1.3) depends on the diffusion constant $\delta > 0$ and resources $m \in \mathcal{M}$. We denote the total population at stable equilibrium as $U = U(m, \delta)$; i.e.

$$U(m, \delta) := \sum_{i=1}^{N} u_i,$$

with $(u_1, \ldots, u_N)$ being the positive steady states from (1.1).

Given $m > 0$ and $\delta > 0$, our goal is to find a vector $m = (m_1, \ldots, m_N)$ satisfying (1.2) to maximize $U(m, \delta)$. If we regard the optimal distribution m as a function on the discrete set $\Omega = \{1, 2, \ldots, N\}$, we might first suspect that m is an indicator function with weight, i.e. $m = a\chi_E$ for some $E \subset \Omega$ and constant $a > 0$. Here χ_E denotes the indicator function on set E. The main purpose of this paper is to characterize such set and weight, and to reveal the complexity in characterizing these optimal distributions.

Our first main result is stated as follows:

Theorem 1 (Global maximizer for large δ) *Define*

$$m_1^* = (0, 0, 0, \ldots, m) \quad \textit{and} \quad m_2^* = (m, 0, 0, \ldots, 0).$$

Then there exists a positive constant $\Delta_{N,m} > 0$ such that for any $\delta > \Delta_{N,m}$ and for any $\tilde{m} \in \mathcal{M} \setminus \{m_1^*, m_2^*\}$, the total population satisfies $U(m_1^*, \delta) = U(m_2^*, \delta) > U(\tilde{m}, \delta)$.

Note that this theorem is consistent with the result of Mazari et al. (2020) for (1.4) claiming that large diffusion tends to well mix the populations. Biologically, Theorem 1 suggests that it is advantageous to concentrate the resources in a single patch in order to maximize the total population in well mixed populations.

If we decrease the diffusion rate, the habitats will become less mixed. How should the resources be distributed in poorly mixed habitats to maximize the total population? The next theorem shows that the global maximizer for sufficiently small δ is fragmented, and there are some specific guiding rules of fragmentation in the multi-patch model (1.3). Interestingly, these guiding rules look different for the cases

$N=3p, 3p+1, 3p+2$, where p is any positive integer. In this connection, for any given $N\geq 3$, we set

$$p := \left\lfloor \frac{N}{3} \right\rfloor, \qquad r := \left\lfloor \frac{p+1}{2} \right\rfloor, \tag{2.1}$$

where $\lfloor x \rfloor$ denotes the floor function of real number x.

Theorem 2 (Global maximizer for small δ) *Given any $N\geq 3$, let p, r be positive integers given by (2.1). For any $m>0$, define $P_m=(0, m/p, 0)$, $P_{m_*}=(0, m_*, 0)$, $m^*=\big(0, m^*, 0, m^*, 0\big)$, where*

$$m^* = \frac{(1+\sqrt{2})^2 m}{2\{4(p-1)+(1+\sqrt{2})^2\}}, \quad m_* = \frac{4m}{4(p-1)+(1+\sqrt{2})^2}. \tag{2.2}$$

Choose $\eta \in (0, p^*]$ arbitrarily, where

$$p^* := \begin{cases} m/p & \text{if } N=3p, 3p+1, \\ m^* & \text{if } N=3p+2. \end{cases}$$

Define a set

$$\mathcal{M}_\eta := \{m \in \mathcal{M} \mid m_i \geq \eta \text{ or } m_i = 0 \text{ for all } i \in \Omega\}.$$

Then there exist positive constant $\delta_{N,m,\eta}>0$ and $m\in\mathcal{M}_\eta$ such that $U(m,\delta) > U(\tilde{m},\delta)$ holds for any $\delta \in (0, \delta_{N,m,\eta})$ and any $\tilde{m}\in\mathcal{M}_\eta$. Furthermore, the optimal resource distribution m is explicitly given as follows:

(i) *If $N=3p$ and $p\geq 1$, then*

$$m = (\underbrace{P_m, P_m, \ldots, P_m}_{p}).$$

(ii) *If $N=3p+1$ and $p\geq 1$, there are two cases:*

(a) *For $p=2r$ and $r\geq 1$, then*

$$m = (\underbrace{P_m, \ldots, P_m}_{r}, \overset{3r+1}{\check{0}}, \underbrace{P_m, \ldots, P_m}_{r}).$$

(b) *For $p=2r-1$ and $r\geq 1$, then*

$$m = (\underbrace{P_m, \ldots, P_m}_{r}, \overset{3r+1}{\check{0}}, \underbrace{P_m, \ldots, P_m}_{r-1}),$$

or

$$m = (\underbrace{P_m, \ldots, P_m}_{r-1}, \overset{3(r-1)+1}{\check{0}}, \underbrace{P_m, \ldots, P_m}_{r}).$$

(iii) *If $N = 3p + 2$ and $p \geq 1$, there are two cases:*

(a) *For $p = 2r$ and $r \geq 1$, then*

$$m = (\underbrace{P_{m_*}, \ldots, P_{m_*}}_{r-1}, m^*, \underbrace{P_{m_*}, \ldots, P_{m_*}}_{r}),$$

or

$$m = (\underbrace{P_{m_*}, \ldots, P_{m_*}}_{r}, m^*, \underbrace{P_{m_*}, \ldots, P_{m_*}}_{r-1}).$$

(b) *For $p = 2r - 1$ and $r \geq 1$, then*

$$m = (\underbrace{P_{m_*}, \ldots, P_{m_*}}_{r-1}, m^*, \underbrace{P_{m_*}, \ldots, P_{m_*}}_{r-1}).$$

If we regard the patches in model (1.1) as the vertices of a connected graph, such graphs are the least connected ones among all path connected graphs with N-vertices, so that the population in (1.1) is less connected in comparison to other graphs. Small diffusion rate will further weaken the mixing of the population. Biologically, Theorem 2 suggests that in order to maximize the total population in weakly connected and poorly mixed habitats, it is advantageous to distribute the resources in fragmented manners. This is in strong contrast with Theorem 1 for large diffusion rate.

Remark 1 Let $m \in \mathcal{M}$ be given as in Theorem 2. Choose $\tilde{m} \in \mathcal{M} \setminus \{m\}$ arbitrarily. The proof of Theorem 2 implies that there exists some positive number $\delta_{\tilde{m}}$ such that for any $\delta \in (0, \delta_{\tilde{m}})$, it follows that $U(m, \delta) > U(\tilde{m}, \delta)$. Hence, given any $\tilde{m} \in \mathcal{M} \setminus \{m\}$, it is not a global maximizer for sufficiently small δ. We suspect that such $\delta_{\tilde{m}}$ can be chosen independently of $\tilde{m} \in \mathcal{M} \setminus \{m\}$, that is, there exists some $\delta = \delta(N, m) > 0$ such that for any $\delta \in (0, \delta(N, m))$, $U(m, \delta) > U(\tilde{m}, \delta)$ holds for all $\tilde{m} \in \mathcal{M} \setminus \{m\}$.

Remark 2 It will be interesting to study the switch of the optimal distributions between small and large δ. Take $N = 4$ as an example: For small δ, the optimal resource distribution is concentrated in either patch #2 or #3 (called type 1). In contrast, for large δ, it is concentrated in either patch #1 or #4 (called type 2). For intermediate δ, there might be multiple local maximizers, and it is difficult to determine which local maximizers are the global maximizers. It is possible that there exists a sudden switch of the global maximizers from type 1 to type 2, i.e. there exists some $\delta^* > 0$ such that the global maximizers are of type 1 for $\delta < \delta^*$, and they are of type 2 for $\delta > \delta^*$. Determining the global and local maximizers for general diffusion rate is a challenging question.

As mentioned earlier, there are some general guiding rules of fragmentation in the multi-patch model (1.3), as specified by Theorem 2. In the following we use some graphs to illustrate these guiding rules for $N = 3p, 3p+1, 3p+2$, respectively. It turns out that there is a unified guiding rule for arbitrary N.

For $N = 3, 4, 5$, Theorem 2 implies that the optimal resource distributions are given by, respectively,

$$m = \begin{cases} (0, m, 0), & N = 3; \\ (0, m, 0, 0) \text{ or } (0, 0, m, 0) & N = 4; \\ \left(0, \frac{m}{2}, 0, \frac{m}{2}, 0\right), & N = 5. \end{cases}$$

For general $N \geq 3$, Theorem 2(i)-(ii) imply that for $N = 3p, 3p+1$, $m = (m/p)\chi_E$ for some set $E \subset \Omega$. Interestingly, the optimal distribution for $N = 3p+2$ is the sum of two indicator functions, i.e. $m = m^*\chi_{E_1} + m_*\chi_{E_2}$ for two sets $E_i \subset \Omega$, $i = 1, 2$, with m_*, m^* given in (2.2). The main contribution of Theorem 2 is to characterize these sets and corresponding weights.

1. For $N = 3p$, $E = \{2, 5, 8, \ldots, 3p-1\}$ and the corresponding weight is m/p, as illustrated in Fig. 1 for cases $N = 3, 6, 9$, respectively.
2. For $N = 3p+1$, while the weight remains to be m/p as the case $N = 3p$, the set E appears to be more complicated. In Fig. 2, we start with 4-patch optimal distribution $(0, m, 0, 0)$, and add three new patches to its right to obtain the optimal distribution for $N = 7$ with $E = \{2, 6\}$ and weight $m/2$. Then we add three new patches to the left of 7-patch to obtain the optimal distribution for $N = 10$, with the patches renumbered from left to right, so that $E = \{2, 5, 9\}$ and weight $m/3$. We can repeat this process for $N = 3p+1$ for all p. Similarly, we can start with the other optimal distribution $(0, 0, m, 0)$ for $N = 4$ and repeat the same process (but switch the order of left and right) to obtain the rest of optimal resource distributions for $N = 3p+1$. For the sake of brevity, we do not include the second scenario in Fig. 2.
3. For $N = 3p+2$, we start with 5-patch optimal distribution in Fig. 3, and add three new patches to its right to obtain the optimal distribution for $N = 8$ with $E_1 = \{2, 4\}$ and weight m^*, $E_2 = \{7\}$ and weight m_*. Then we add three new patches to the left of 8-patch to obtain the optimal distribution for $N = 11$, with $E_1 = \{5, 7\}$ and weight m^*, $E_2 = \{2, 10\}$ and weight m_*. We can repeat this process for $N = 3p+2$ for all p. Similarly, we can start with the same optimal distribution for $N = 5$ and repeat the process but switching the order of left and right to obtain the rest of optimal resource distributions for $N = 3p+2$. For the sake of brevity, we do not include the second scenario in Fig. 3.

In summary, if we start from the cases $N = 3, 4, 5$, at each step add three new patches, alternatively to right and left (or switch the order to left and right), then we can obtain the optimal resource distributions for arbitrary N. Note that for $N = 3p+1$, the choice of the left-right order depends on the initial optimal distribution for $N = 4$.

Theorem 2 refers to a global-maximizer when there is minimum amount of resources $\eta > 0$ for each patch. When $\eta = 0$, it is very difficult to determine a

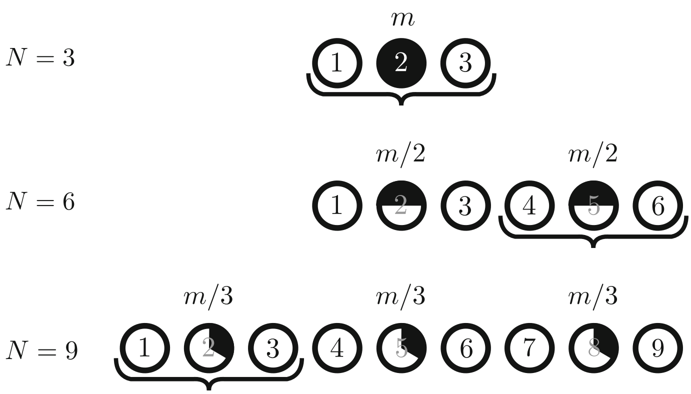

Fig. 1 Illustrations of the optimal resource distributions for $N = 3, 6, 9$, from which we can observe the general pattern for $N = 3p$, as given in Theorem 2(i). The patches with underbraces are newly added ones. At each step we renumber the patches from left to right

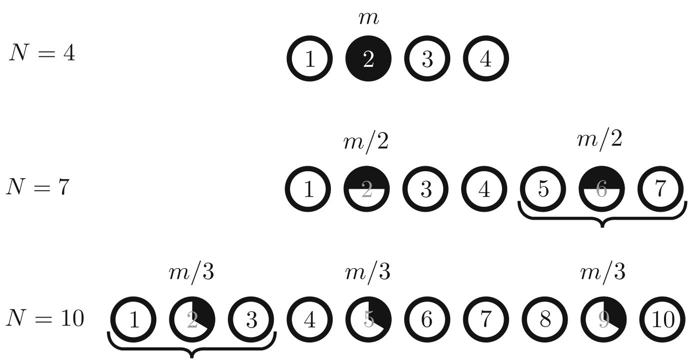

Fig. 2 Illustrations of the optimal resource distributions for the cases $N = 4, 7, 10$, starting with the optimal distribution $(0, m, 0, 0)$ for $N = 4$, from which we can see the guiding rule in determining the optimal distributions for $N = 3p + 1$, as given by Theorem 2(ii). For brevity, we do not include the other scenario, i.e. starting with the optimal distribution $(0, 0, m, 0)$ for $N = 4$ and repeating the same process (but switch the order of left and right)

global maximizer. However, in the two patch case, we obtain the global maximizers for all $\delta > 0$ in the following result:

Theorem 3 (Global maximizer for two patch) *In the case $N = 2$, define*

$$m_1^* = (0, m) \quad \text{and} \quad m_2^* = (m, 0).$$

Then for any $m \in \mathcal{M} \setminus \{m_1^*, m_2^*\}$, it follows that $U(m_1^*, \delta) = U(m_2^*, \delta) > U(m, \delta)$ for all $\delta > 0$.

Two-patch habitat is well connected. Biologically, Theorem 3 suggests that for well connected habitats, it could be advantageous to concentrate the resources in a single patch in order to maximize the total population. This also echoes the conclusions of Theorem 1 for well mixed habitats.

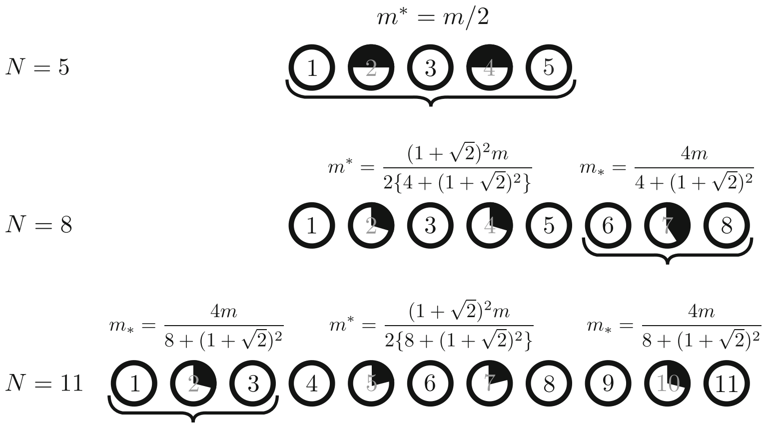

Fig. 3 Illustrations of the optimal resource distributions for the cases $N = 5, 8, 11$. from which we can see the guiding rule in determining the optimal distributions for $N = 3p + 2$, as given by Theorem 2(iii). For brevity, we do not include the other scenario, i.e. starting with the same optimal distribution for $N = 5$ but switching the order of left and right

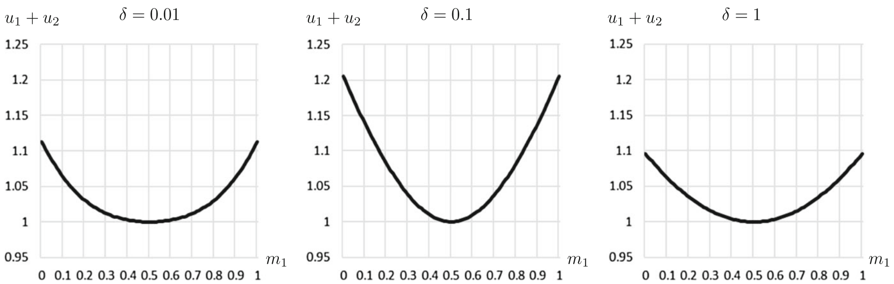

Fig. 4 Graphs of the total population as function of $m_1 \in [0, 1]$ for the two patch case, $m = 1$, $\delta = 0.01, 0.1, 1$, respectively

In Fig. 4, we assume $m_1 + m_2 = 1$ and $m_1, m_2 \geq 0$. For $\delta = 0.01, 0.1, 1$, numerical simulations illustrate that $u_1 + u_2$, as a function of m_1, attains the maximum value at $m_1 = 0, 1$, the minimum value at $m_1 = 1/2$, and there is no other critical point. Hence, $u_1 + u_2$ is decreasing in $m_1 \in (0, 1/2)$ and it is symmetric with respect to $m_1 = 1/2$.

This paper is organized as follows. In Sect. 3, we establish some basic properties of the population density in each patch. In Sect. 4, we calculate the global maximizer of the total population with large diffusion constant and prove Theorem 1. In Sect. 5, we first calculate the first variation of U with respect to m, and then demonstrate Theorem 3. In Sect. 6, we consider the case $N \geq 3$ and restrict the candidates of the global maximizer, and give the proof for $N = 3p, 4, 5$. In Sect. 7, we give the complete proof of Theorem 2. In Sects. 7.2 and 7.3, we treat the cases $N = 3p + 1$ and $N = 3p + 2$, respectively. In Sect. 8 we discuss the main findings in this paper. Some technical lemmas are postponed to the Appendices.

3 Preliminaries

In this section, we choose $N \geq 2$ and $m \in \mathcal{M}$ arbitrarily. We define the degree of u_i as $\delta \downarrow 0$ by

$$\underline{\deg}\, u_i := \sup \left\{ \gamma \in \mathbb{R} \mid \limsup_{\delta \downarrow 0} \left| \frac{u_i}{\delta^\gamma} \right| < \infty \right\}.$$

We also define the order of u_i as $\delta \to \infty$ by

$$\overline{\deg}\, u_i := \sup \left\{ \gamma \in \mathbb{R} \mid \limsup_{\varepsilon \downarrow 0} \left| \frac{u_i}{\varepsilon^\gamma} \right| < \infty \right\},$$

where $\varepsilon := 1/\delta$. The degree and order of u_i are used to calculate the leading term of u_i when we expand u_i with respect to small and large δ, respectively.

We give two basic results to prove our main results.

Lemma 1 *A non-trivial solution of* (1.3) *exists and satisfies $u_i > 0$ for all $i \in \Omega$. Moreover, it is a globally stable in* (1.1).

Proof The main idea of this proof is due to Takeuchi (1989). Let J_0 denote the Jacobian matrix at the trivial equilibrium point 0 for (1.1). We use Rayleigh quotient to have

$$\begin{aligned} \sup_{x \in \mathbb{R}^N \setminus \{0\}} \frac{{}^t x J_0 x}{\|x\|} &= \sup_{x \in \mathbb{R}^N \setminus \{0\}} \frac{\left\{ -\delta \sum_{i=1}^{N-1} (x_{i+1} - x_i)^2 + \sum_{i=1}^{N} m_i x_i^2 \right\}}{\|x\|^2} \\ &\geq \sum_{i=1}^{N} m_i = m > 0. \end{aligned}$$

Therefore, the trivial equilibrium point is unstable for all $\delta > 0$ and the unique solution $v_i(t)$ is strictly positive for all $t \in \mathbb{R}_+$ and $\delta > 0$. Moreover, we use (1.3) to obtain unique positive non-trivial equilibrium point as follows:

$$u_1 = \frac{m_1 - \delta}{2} + \left(\frac{(m_1 - \delta)^2}{4} + \delta u_2 \right)^{1/2}, \tag{3.1}$$

$$u_i = \frac{m_i - 2\delta}{2} + \left(\frac{(m_i - 2\delta)^2}{4} + \delta (u_{i-1} + u_{i+1}) \right)^{1/2}, \tag{3.2}$$

$$u_N = \frac{m_N - \delta}{2} + \left(\frac{(m_N - \delta)^2}{4} + \delta u_{N-1} \right)^{1/2}. \tag{3.3}$$

Global stability of this equilibrium point is in the same manner as Takeuchi (1989), so we omit the proof. □

Next, we show that a global minimizer of the total population is given as follows. We use this result in Sect. 4.

Proposition 1 (Global minimizer) *Let $\{m_i\}_{i\in\Omega} \in \mathcal{M}$ be given by $m_i = m/N$ for all $i \in \Omega$. Then $\{m_i\}$ is a unique global minimizer of the total population at equilibrium for all δ.*

Proof We divide (1.3) by u_i for each $i \in N$ to get

$$m_i - u_i + \delta\left(\frac{u_{i-1}}{u_i} + \frac{u_{i+1}}{u_i} - 2\right) = 0. \tag{3.4}$$

Summing up (3.4) in $i \in \Omega$, by the definition of U and (1.2) we have

$$\begin{aligned} &U(m,\delta) - m \\ &\quad = \delta\left\{\left(\frac{u_2}{u_1} + \frac{u_1}{u_2}\right) + \left(\frac{u_3}{u_2} + \frac{u_2}{u_3}\right) + \cdots + \left(\frac{u_N}{u_{N-1}} + \frac{u_{N-1}}{u_N}\right) - 2(N-1)\right\} \\ &\quad \geq \delta(2(N-1) - 2(N-1)) = 0, \end{aligned} \tag{3.5}$$

where $u_0 = u_1$ and $u_N = u_{N+1}$ are also used. Here the equality holds if and only if $u_i \equiv$ Constant. Hence we obtain $m_i \equiv m/N$, where $m > 0$ is the lower bound of the total population. □

Remark 3 It is known that a global minimizer of the total population of (1.4) is constant; see, e.g. Lou (2006). It still holds when we use the multi-patch model (1.1), that is, $m_i \equiv m/N$ for all $i \in \Omega$ is the global minimizer for all $\delta > 0$. Since we cannot locate the proof of the global minimizer in the patchy environment, we include a proof here.

4 Global maximizer for large δ: Proof of Theorem 1

In this section, we choose $m \in \mathcal{M}$ arbitrarily. We first show that an analogy with Theorem A.2 in Cantrell et al. (1996) holds in any patchy environment. Recall that $\varepsilon = 1/\delta$ as in Sect. 3.

Lemma 2 *For every $m \in \mathcal{M}$, the solution of (1.3) satisfies $u_i = c_0 + o(1)$ as $\varepsilon \to 0$ for all $i \in \Omega$, where c_0 is a constant independent of $i \in \Omega$.*

Proof From the argument of Lemma 1, we have $u_i > 0$ for all $\delta > 0$. By the maximum principle, it is easy to show that

$$\max\{u_i \mid i \in \Omega\} \leq \max\{m_i \mid i \in \Omega\} \leq m \tag{4.1}$$

for all $\delta > 0$. Hence $\overline{\deg}\, u_i \geq 0$.

Next, we use (1.3) to have

$$\begin{cases} u_2/u_1 = 1 + \varepsilon(u_1 - m_1), \\ (u_{i-1} + u_{i+1})/u_i = 2 + \varepsilon(u_i - m_i), \\ u_{N-1}/u_N = 1 + \varepsilon(u_N - m_N). \end{cases} \tag{4.2}$$

This means that $\overline{\deg}\, u_i \equiv \gamma \geq 0$ for all $i \in \Omega$ with some constant γ. Assume that $\gamma > 0$. Then we can rewrite u_i as

$$u_i := u_{i,\gamma}\varepsilon^{\gamma} + o(\varepsilon^{\gamma}),$$

where $u_{i,\gamma}$ is a coefficient of u_i of order ε^{γ}. Then we have

$$U(m,\varepsilon) = \sum_{i=1}^{N} u_{i,\gamma}\varepsilon^{\gamma} + o(\varepsilon^{\gamma}) \to 0 \text{ as } \varepsilon \to 0.$$

However, this contradicts Proposition 1. From (4.2), it follows that $u_{i,0} \equiv c_0$ for all $i \in \Omega$. $\square$

Let us complete the proof of Theorem 1. We define $\{v_i\}_{i\in\Omega}$ as

$$v_i := \frac{u_i - c_0}{\varepsilon}, \qquad i = 1, 2, \dots, N.$$

From (1.3), we have

$$\begin{cases} u_1(m_1 - u_1) + v_2 - v_1 = 0, \\ u_i(m_i - u_i) + v_{i-1} + v_{i+1} - v_i = 0, \\ u_N(m_N - u_N) + v_{N-1} - v_N = 0. \end{cases} \tag{4.3}$$

In fact, $|v_i - v_j|$ is uniformly bounded for all $i, j \in \Omega$ since $|u_i(m_i - u_i)| \leq m^2$ for all $i \in \Omega$, which is clear from (4.1). Then we can express $v_i = \varepsilon^{\gamma}\tilde{v}_i$, where $\tilde{v}_i$ is bounded for sufficiently small ε. Note that $\gamma > -1$ in view of (4.1). To compute c_0 explicitly, we rewrite (4.3) as

$$\begin{cases} m_1(c_0 + \varepsilon^{1+\gamma}\tilde{v}_1) - (c_0 + \varepsilon^{1+\gamma}\tilde{v}_1)^2 + \varepsilon^{\gamma}(\tilde{v}_2 - \tilde{v}_1) = 0, \\ m_i(c_0 + \varepsilon^{1+\gamma}\tilde{v}_i) - (c_0 + \varepsilon^{1+\gamma}\tilde{v}_i)^2 + \varepsilon^{\gamma}(\tilde{v}_{i-1} + \tilde{v}_{i+1} - 2\tilde{v}_i) = 0, \\ m_N(c_0 + \varepsilon^{1+\gamma}\tilde{v}_N) - (c_0 + \varepsilon^{1+\gamma}\tilde{v}_N)^2 + \varepsilon^{\gamma}(\tilde{v}_{N-1} - \tilde{v}_N) = 0. \end{cases} \tag{4.4}$$

Adding both sides of these equations, we obtain

$$c_0 m - Nc_0^2 + \varepsilon^{1+\gamma}\sum_{i=1}^{N}(m_i\tilde{v}_i - 2c_0\tilde{v}_i - \varepsilon^{1+\gamma}\tilde{v}_i^2) = 0.$$

Since $1 + \gamma > 0$, we have $c_0 = m/N$. Further, we claim $\gamma = 0$. Suppose that $-1 < \gamma < 0$. From (4.4), we have $\tilde{v}_i \equiv$ Constant for all $i \in \Omega$. This means that u_i is also constant and m_i is given by $m_i \equiv m/N$. Indeed, this is a global minimizer of $U(m,\varepsilon)$ by (3.5). Similarly, if $\gamma > 0$, then we have $m_i \equiv m/N$, which is again the global minimizer.

From the above argument, we have $\overline{\deg}\, v_i \equiv 0$. Therefore (4.4) must be written as

$$Av = -c_0({}^t m - c_0) + O(\varepsilon^2), \tag{4.5}$$

where

$$A = \begin{pmatrix} m_1^{(c)}\varepsilon - 1 & 1 & & & & 0 \\ 1 & m_2^{(c)}\varepsilon - 2 & 1 & & & \\ & 1 & m_3^{(c)}\varepsilon - 2 & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & m_{N-1}^{(c)}\varepsilon - 2 & 1 \\ 0 & & & & 1 & m_N^{(c)}\varepsilon - 1 \end{pmatrix},$$

$m = (m_1, m_2, \ldots, m_N)$, $m_i^{(c)} := m_i - 2c_0$, and $v = {}^t(v_1, v_2, \ldots, v_N)$. Then we can calculate $\det A$ as

$$\det A = (-1)^{N+1} \sum_{i=1}^{N} m_i^{(c)}\varepsilon + O(\varepsilon^2).$$

Let $\tilde{a}_{i,j}$ denote the (i, j)-cofactor of A, which can be expressed as

$$\begin{cases} \tilde{a}_{i,j} = (-1)^{N+1}\left[1 - \left\{\sum_{l=1}^{N-j}(N - j + 1 - l)m_{N+1-l}^{(c)} + \sum_{l=1}^{i-1}(i - l)m_l^{(c)}\right\}\varepsilon\right] \\ \qquad +O(\varepsilon^2) \quad \text{if } i \le j, \\ \tilde{a}_{i,j} = (-1)^{N+1}\left[1 - \left\{\sum_{l=1}^{N-i}(N - i + 1 - l)m_{N+1-l}^{(c)} + \sum_{l=1}^{i-1}(j - l)m_l^{(c)}\right\}\varepsilon\right] \\ \qquad +O(\varepsilon^2) \quad \text{if } i \ge j. \end{cases}$$

Multiplying A^{-1} to (4.5) and using these cofactors, we have

$$\sum_{i=1}^{N} v_i = -(\det A)^{-1} \sum_{i=1}^{N} \left(c_0(m_i - c_0) \sum_{j=1}^{N} \tilde{a}_{i,j} \right).$$

Here we compute

$$
\begin{aligned}
&\sum_{i=1}^{N}\left(c_0(m_i-c_0)\sum_{j=1}^{N}\tilde{a}_{i,j}\right)\\
&\quad=(-1)^{N+1}\frac{c_0}{2}\Bigg[2N\sum_{i=1}^{N-1}m_i\left(\sum_{k=1}^{N-1}km_{i+k}\right)\\
&\qquad+c_0\sum_{i=1}^{N-1}\left(m_i\sum_{k=1}^{N-i}\{-k^2+(-2(N+i)+1)k\}\right)\\
&\qquad+c_0\sum_{i=1}^{N-1}\sum_{k=1}^{N-i}km_{i+k}(2i-1+k-4N)+4Nc_0^2\sum_{i=1}^{N-1}\sum_{k=1}^{N-i}k\Bigg]\varepsilon+O(\varepsilon^2).
\end{aligned}
$$

By $c_0=m/N$, we obtain

$$
\begin{aligned}
\sum_{i=1}^{N}v_i&=\frac{m}{N}\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)^2m_i-\sum_{1\le i<j\le N}(j-i)m_im_j\\
&\quad+\frac{m^2}{12N}(N^2-1)+O(\varepsilon).
\end{aligned}
\tag{4.6}
$$

Now let us maximize the right-hand side of (4.6). We compute

$$
\begin{aligned}
&\frac{m}{N}\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)^2m_i-\sum_{1\le i<j\le N}(j-i)m_im_j\\
&\quad\le\sum_{i=1}^{\lfloor N/2\rfloor}\left\{\frac{m}{N}\left(i-\frac{N+1}{2}\right)^2(m_i+m_{N+1-i})-(N+1-2i)m_im_{N+1-i}\right\}\\
&\quad\le\frac{m}{N}\left(\frac{N-1}{2}\right)^2m,
\end{aligned}
$$

where the equality holds if and only if $m=(m,0,\ldots,0)$ or $(0,\ldots,0,m)$. We can conclude Theorem 1 proved. □

5 Asymptotic expansion with respect to *m*: Proof of Theorem 3

In this section, we show Theorem 3. To this purpose, we prepare expansion of $U(m,\delta)$ with respect to m for all $\delta>0$.

Choose $\delta>0$ arbitrarily. We set $m:=m^0+\varepsilon g$, where $\varepsilon>0$ is a small parameter, and compute the first variations of $U(m,\delta)$ with respect to m. We indicate the dependence of the positive solution $u:=(u_1,\ldots,u_N)$ to (1.3) on $m\in\mathcal{M}$ only by writing $u=u(m)$.

Proposition 2 *For every $m^0 \in \mathcal{M}$, choose $\varepsilon > 0$ and $g := (g_1, \ldots, g_N) \in \mathbb{R}^N$ such that $m := m^0 + \varepsilon g$ with $m \in \mathcal{M}$. Then*

$$U(m^0 + \varepsilon g, \delta) = \sum_{i=1}^{N} u_i^0 + \varepsilon \sum_{i=1}^{N} u_i^1 + o(\varepsilon) \quad \text{as } \varepsilon \to 0,$$

where $u^0 := (u_1^0 \ldots, u_N^0)$ and $u^1 := (u_1^1 \ldots, u_N^1)$ are given by

$$\begin{cases} (m_i^0 - u_i^0)u_i^0 + \delta(u_{i-1}^0 + u_{i+1}^0 - 2u_i^0) = 0 \quad \text{in } \Omega, \\ u_0^0 = u_1^0, \ u_{N+1}^0 = u_N^0, \end{cases} \tag{5.1}$$

and

$$(D_\delta + \mathrm{diag}(m^0 - 2u^0))u^1 = -\,\mathrm{diag}(g)u^0, \tag{5.2}$$

respectively. Here

$$D_\delta = \begin{pmatrix} -\delta & \delta & & & & 0 \\ \delta & -2\delta & \delta & & & \\ & \delta & -2\delta & \ddots & & \\ & & \ddots & \ddots & \ddots & \\ & & & \ddots & -2\delta & \delta \\ 0 & & & & \delta & -\delta \end{pmatrix}.$$

Moreover, the linear operator $D_\delta + \mathrm{diag}(m^0 - 2u^0)$ is invertible.

Proof We first show that the linear operator $D_\delta + \mathrm{diag}(m^0 - 2u^0)$ is invertible. We use properties of the principal eigenvalue of the eigenvalue problem

$$\lambda \Phi = (D_\delta + \mathrm{diag}\, q)\Phi, \tag{5.3}$$

where $q \in \mathbb{R}^N$. It is well known that (5.3) has a maximum eigenvalue $\lambda_0(q)$, which is characterized as

$$\lambda_0(q) = \sup_{x \in \mathbb{R}^N \setminus \{0\}} \frac{\left\{-\delta \sum_{i=1}^{N-1}(x_{i+1} - x_i)^2 + \sum_{i=1}^{N} q_i x_i^2\right\}}{\|x\|^2},$$

by the Rayleigh quotient.

Let $k > 0$ be a sufficiently large constant such that $D_\delta + \mathrm{diag}(m^0 - u^0 + k)$ is an irreducible matrix. By the Perron-Frobenius theorem, $\lambda_0(m^0 - u^0 + k) = k$ is the Perron-Frobenius eigenvalue since u^0 is the associated eigenvector. Then we have $\lambda_0(m^0 - u^0) = 0$. Hence we get

$$\lambda_0(m^0 - 2u^0) < \lambda_0(m^0 - u^0) = 0. \tag{5.4}$$

The second half of the proof is devoted to the asymptotic expansion of $U(m,\delta)$. The main idea of this proof is due to (Ding et al. (2010)). Let $\varepsilon > 0$ be a sufficiently small constant. We shall show below that there exists some constant $C_1 > 0$ such that

$$\left\| \frac{u-u^0}{\varepsilon} \right\| \le C_1, \tag{5.5}$$

where $u := u(m^0 + \varepsilon g)$.

To prove (5.5), we substitute $m = m^0 + \varepsilon g$ for (1.3) to obtain

$$u_i((m_i^0 + \varepsilon g_i) - u_i) + \delta(u_{i-1} + u_{i+1} - 2u_i) = 0. \tag{5.6}$$

Subtracting (5.6) from the first equation in (5.1) and dividing by ε, we have

$$\left(\frac{u_{i-1}-u_{i-1}^0}{\varepsilon} + \frac{u_{i+1}-u_{i+1}^0}{\varepsilon} - 2\frac{u_i-u_i^0}{\varepsilon} \right) + (m_i^0 - (u_i + u_i^0)) \left(\frac{u_i-u_i^0}{\varepsilon} \right) + g_i u_i^0 = 0.$$

Multiplying both sides of the above equality by $(u_i - u_i^0)/\varepsilon$ and adding $i = 1$ to N, we have

$$\sum_{i=1}^{N-1} \left| \frac{u_{i+1}-u_{i+1}^0}{\varepsilon} - \frac{u_i-u_i^0}{\varepsilon} \right|^2 - \sum_{i=1}^{N} (m_i^0 - (u_i + u_i^0)) \left(\frac{u_i-u_i^0}{\varepsilon} \right)^2 = \sum_{i=1}^{N} g_i u_i \left(\frac{u_i-u_i^0}{\varepsilon} \right). \tag{5.7}$$

Now, we give the upper bound in (5.5). By (4.1), we have $\sup_{i\in\Omega} m_i^0 - (u_i + u_i^0) \le 3m$. Then

$$-\lambda_0(m^0 - (u+u^0)) = \inf_{x\in\mathbb{R}^N\setminus\{0\}} \frac{\sum_{i=1}^{N-1} |x_{i+1}-x_i|^2 - \sum_{i=1}^{N} (m_i^0 - (u_i+u_i^0))x_i^2}{\|x\|^2}.$$

Choosing $x = (u-u^0)/\varepsilon$, we have

$$-\lambda_0(m^0 - (u+u^0)) \left\| \frac{u-u^0}{\varepsilon} \right\|^2 \le \sum_{i=1}^{N-1} \left| \frac{u_{i+1}-u_{i+1}^0}{\varepsilon} - \frac{u_i-u_i^0}{\varepsilon} \right|^2 - \sum_{i=1}^{N} (m_i^0 - (u_i + u_i^0)) \left(\frac{u_i-u_i^0}{\varepsilon} \right)^2.$$

By (5.7) and the Cauchy–Schwarz inequality, we compute

$$
\begin{aligned}
&-\lambda_0(m^0-(u+u^0))\left\|\frac{u-u^0}{\varepsilon}\right\|^2\\
&\le \sum_{i=1}^N g_i u_i\left(\frac{u_i-u_i^0}{\varepsilon}\right)\le \sup_{i\in\Omega} g_i\,\|u\|\left\|\frac{u-u^0}{\varepsilon}\right\|.
\end{aligned}
$$

Note that we have $\lambda_0(m^0-(u+u^0))\to\lambda_0(m^0-2u^0)$ as $\varepsilon\to 0$. From (5.4), there exists $\rho_\delta>0$ such that $-\lambda_0(m_0-(u+u_0))>\rho_\delta>0$ for all ε sufficiently small. Thus, we obtain

$$
\left\|\frac{u-u_0}{\varepsilon}\right\|^2\le\frac{1}{\rho_\delta}\left(\sup_{i\in\Omega} g_i\,\|u\|\left\|\frac{u-u^0}{\varepsilon}\right\|\right).
$$

This proves (5.5). □

To prove Theorem 3, we consider a critical point of $U(m,\delta)$. By (5.2) we have

$$
\begin{aligned}
\sum_{i=1}^N g_i &= -\sum_{i=1}^N\left(\frac{1}{u_i^0}\delta(u_{i-1}^1+u_{i+1}^1-2u_i^1)+\left(\frac{m_i^0}{u_i^0}-2\right)u_i^1\right)\\
&=-\sum_{i=1}^N\left(\delta\left(\frac{1}{u_{i-1}^0}+\frac{1}{u_{i+1}^0}-2\frac{1}{u_i^0}\right)+\left(\frac{m_i^0}{u_i^0}-2\right)\right)u_i^1,
\end{aligned}
$$

where g satisfies the condition of Proposition 2. Here, the integrands of the right-hand side of the above equality can be calculated as

$$
\begin{aligned}
&\delta\left(\frac{1}{u_{i-1}^0}+\frac{1}{u_{i+1}^0}-2\frac{1}{u_i^0}\right)+\left(\frac{m_i^0}{u_i^0}-2\right)\\
&=\delta\left(\frac{1}{u_{i-1}^0}+\frac{1}{u_{i+1}^0}-2\frac{1}{u_i^0}\right)+\frac{u_i^0(m_i^0-u_i^0)}{(u_i^0)^2}-1\\
&=-\delta\left\{\frac{u_{i-1}^0+u_{i+1}^0-2u_i^0}{(u_i^0)^2}-\left(\frac{1}{u_{i-1}^0}+\frac{1}{u_{i+1}^0}-2\frac{1}{u_i^0}\right)\right\}-1.
\end{aligned}
$$

Then we obtain

$$
\sum_{i=1}^N g_i=\sum_{i=1}^N\left(\delta\left\{\frac{u_{i-1}^0+u_{i+1}^0-2u_i^0}{(u_i^0)^2}-\left(\frac{1}{u_{i-1}^0}+\frac{1}{u_{i+1}^0}-2\frac{1}{u_i^0}\right)\right\}+1\right)u_i^1.
$$

As mentioned above, the linear operator of (5.2) is invertible. Suppose that $\phi \in \mathbb{R}^N$ and define g by

$$g := -\operatorname{diag}\left(\frac{1}{u_1^0}, \dots, \frac{1}{u_N^0}\right)(D_\delta + \operatorname{diag}(m^0 - 2u^0))\phi.$$

Then u^1 is equal to ϕ, that is

$$(D_\delta + \operatorname{diag}(m^0 - 2u^0))\phi = -\operatorname{diag}(g)u^0.$$

Therefore we can define a function $I[\phi]$ by

$$I[\phi] := \sum_{i=1}^{N} g_i = \sum_{i=1}^{N} \left(\delta w_i(u^0) + 1\right)\phi_i,$$

where

$$w_i(u^0) := \frac{u_{i-1}^0 + u_{i+1}^0 - 2u_i^0}{(u_i^0)^2} - \left(\frac{1}{u_{i-1}^0} + \frac{1}{u_{i+1}^0} - 2\frac{1}{u_i^0}\right).$$

Proof of Theorem 3 Suppose that $N = 2$. Choose $m^0 \in \mathcal{M} \setminus \{(m, 0), (0, m)\}$ arbitrarily. Without loss of generality we may assume $m_1^0 < m_2^0$. For this case, it is easy to see $u_1^0 < u_2^0$. Direct calculations lead to

$$w_1(u^0) = \frac{(u_2^0 - u_1^0)(u_2^0 + u_1^0)}{u_2^0(u_1^0)^2} > 0,$$
$$w_2(u^0) = \frac{(u_1^0 - u_2^0)(u_1^0 + u_2^0)}{u_1^0(u_2^0)^2} < 0.$$

That is, $w_1(u^0) > 0 > w_2(u^0)$.

We also define two constants α, β by

$$\alpha := \delta w_1(u^0) + 1, \quad \beta := \delta w_2(u^0) + 1.$$

Setting ϕ as

$$\phi := \beta e_1 - \alpha e_2,$$

where e_1 and e_2 are canonical basis, we have

$$I[\phi] = \beta\left(\delta w_1(u^0) + 1\right) - \alpha\left(\delta w_2(u^0) + 1\right) = 0.$$

Therefore, $I[\phi] = \sum_{i=1}^{2} g_i = 0$. On the other hand,

$$\sum_{i=1}^{2} u_i^1 = \sum_{i=1}^{2} \phi_i = \beta - \alpha = \delta(w_2(u^0) - w_1(u^0)) \neq 0.$$

This implies that m^0 is not a critical point of $U(m, \delta)$. □

6 Asymptotic expansion of the total population

In this section, we study the expansion of the total population to determine the global maximizer case by case. To prove Theorem 2, we consider the case that δ is sufficiently small. We note that (3.1)–(3.3) can be expanded as

$$u_1 = \begin{cases} m_1 + \left(\dfrac{u_2}{m_1} - 1\right)\delta + o(\delta) & \text{if } m_1 > 0, \\ (\delta u_2)^{1/2} - \dfrac{\delta}{2} + o(\delta) & \text{if } m_1 = 0, \end{cases} \tag{6.1}$$

$$u_i = \begin{cases} m_i + \left(\dfrac{u_{i-1} + u_{i+1}}{m_i} - 2\right)\delta + o(\delta) & \text{if } m_i > 0, \\ (\delta(u_{i-1} + u_{i+1}))^{1/2} - \delta + o(\delta) & \text{if } m_i = 0, \end{cases} \tag{6.2}$$

$$u_N = \begin{cases} m_N + \left(\dfrac{u_{N-1}}{m_N} - 1\right)\delta + o(\delta) & \text{if } m_N > 0, \\ (\delta u_{N-1})^{1/2} - \dfrac{\delta}{2} + o(\delta) & \text{if } m_N = 0, \end{cases} \tag{6.3}$$

which can be obtained by formally expanding (3.1)–(3.3). However, since the rigorous proof is tedious, we will postpone it to Appendix.

Before proceeding to the case of 3 or more patches, we study expansion of the positive solution of (1.3) more precisely. Choose a positive parameter $\eta \in (0, m/N)$ and $m \in \mathcal{M}_\eta$ arbitrarily. We note that expansion of u_i can be expressed as

$$\begin{aligned} u_i = m_i &+ \left(C_{i,1/2}\right)^{1/2} \delta^{1/2} + \left(C_{i,1/4}\right)^{1/2} \delta^{3/4} \\ &+ \cdots + \left(C_{i,1/2^{N-1}}\right)^{1/2} \delta^{(2^{N-1}-1)/2^{N-1}} + C_{i,0}\delta + o(\delta). \end{aligned} \tag{6.4}$$

Here, $C_{i,1/2^k}$, $i \in \Omega$, $k = 1, 2, \ldots, N-1$, and $C_{i,0}$ can be expressed explicitly by the elements of m. Then we can show that there exists $\delta_{N,m,\eta} > 0$ such that for all $\delta \in (0, \delta_{N,m,\eta})$, the total population is expanded as

$$U = m + \sum_{i=1}^{N} \sum_{k=1}^{N-1} \left(C_{i,1/2^k}\right)^{1/2} \delta^{(2^k-1)/2^k} + \sum_{i=1}^{N} C_{i,0}\delta + o(\delta). \tag{6.5}$$

Since the computation is long and tedious, we also postpone it to Appendix. Because the constant term in (6.5) is always equal to m, the first step is to maximize

$\sum_{i=1}^{N}\left(C_{i,1/2}\right)^{1/2}$. In the following subsections, in view of (6.5), we restrict candidates of a global maximizer of the total population.

We define the sets

$$\mathcal{I}_{\eta}(m) := \{i \in \Omega : m_i \geq \eta\}, \quad \mathcal{I}_0(m) := \{i \in \Omega : m_i = 0\}.$$

We also define a subset of $\mathcal{I}_0(m)$ by

$$\mathcal{I}_0^*(m) := \{i \in \mathcal{I}_0(m) : \ m_{i+1} > 0 \text{ or } m_{i-1} > 0\}$$

This subsection will be divided into three parts according to the number of patches. We first consider the case $N = 3p$ with a positive integer p.

Proof (Proof of Theorem 2 (i)) To evaluate $\sum_{i=1}^{N}\left(C_{i,1/2}\right)^{1/2}$ associated with $m \in \mathcal{M}_{\eta}$, the proof is divided into three parts according to $\#\mathcal{I}_{\eta}(m)$, where "#" stands for the number of elements. First, suppose $\#\mathcal{I}_{\eta}(m) < p$. Then $\#\mathcal{I}_0^*(m)$ is at most $2(\#\mathcal{I}_{\eta}(m) - 1)$ so that

$$\begin{aligned}\sum_{i=1}^{N}\left(C_{i,1/2}\right)^{1/2} &\leq \sqrt{2(\#\mathcal{I}_{\eta}(m) - 1)\sum_{i \in \mathcal{I}_0(m)} C_{i,1/2}} \\ &\leq \sqrt{2(\#\mathcal{I}_{\eta}(m) - 1)2m} < \sqrt{2(2p-2)m} < 2\sqrt{p}\sqrt{m}.\end{aligned}$$

Second, suppose $\#\mathcal{I}_{\eta}(m) > p$. Then $\#\mathcal{I}_0^*(m)$ is at most $3p - \#\mathcal{I}_{\eta}(m)$ so that

$$\begin{aligned}\sum_{i=1}^{N}\left(C_{i,1/2}\right)^{1/2} &\leq \sqrt{(3p - \#\mathcal{I}_{\eta}(m))\sum_{i \in \mathcal{I}_0(m)} C_{i,1/2}} \\ &\leq \sqrt{(3p - \#\mathcal{I}_{\eta}(m))2m} \leq \sqrt{2(2p-1)m} < 2\sqrt{p}\sqrt{m}.\end{aligned}$$

Finally, suppose $\#\mathcal{I}_{\eta}(m) = p$. Then $\#\mathcal{I}_0^*(m)$ is also at most $2p$ so that

$$\sum_{i=1}^{N}\left(C_{i,1/2}\right)^{1/2} \leq \sqrt{2p\sum_{i \in \mathcal{I}_0(m)} C_{i,1/2}} \leq \sqrt{2p(2m)} = 2\sqrt{p}\sqrt{m}.$$

Therefore, if the equality holds in the last case for some m, then such m must be a global maximizer.

Now we choose $m \in \mathcal{M}_{\eta}$ as $(\tilde{P}_1, \ldots, \tilde{P}_p)$, where $\tilde{P}_i := (0, m_i, 0)$. Then $\#\mathcal{I}_0^*(m) = 2p$ so that

$$\begin{aligned}\sum_{i=1}^{N}\left(C_{i,1/2}\right)^{1/2} &= 2(\sqrt{m_1} + \sqrt{m_2} + \cdots + \sqrt{m_p}) \\ &\leq 2\sqrt{p(m_1 + m_2 + \cdots + m_p)} = 2\sqrt{p}\sqrt{m},\end{aligned}$$

where the equality holds if and only if $m_1 = m_2 = \cdots = m_p = m/p > 0$. Thus it is shown that $(\tilde{P}_1, \ldots, \tilde{P}_p)$ is a unique global maximizer. $\square$

We second consider the case $N = 3p + 1$. We begin by proving the case $N = 4$.

Proof of Theorem 2 (ii-a). The idea of this proof is the same as that of Theorem 2 (i). First, suppose $\#\mathcal{I}_\eta(m) \geq 3$. Then $\#\mathcal{I}_0(m) \leq 1$ so that

$$\sum_{i=1}^{4} \left(C_{i,1/2}\right)^{1/2} < \sqrt{m} < 2\sqrt{m}.$$

Second, we suppose $\#\mathcal{I}_\eta(m) = 2$. Let $m \in \mathcal{M}_\eta$ be given as $m = (0, m_1, 0, m_2)$ or $m = (m_1, 0, m_2, 0)$ with $m_1 \geq \eta$, $m_2 \geq \eta$ so that

$$\sum_{i=1}^{4} \left(C_{i,1/2}\right)^{1/2} = \sqrt{m_1} + \sqrt{m_1 + m_2} = \sqrt{m_1} + \sqrt{m} < 2\sqrt{m}.$$

The other case is that $m \in \mathcal{M}_\eta$ is given as $m = (0, m_1, m_2, 0)$ or $m = (m_1, 0, 0, m_2)$ with $m_1 \geq \eta$ and $m_2 \geq \eta$ so that

$$\sum_{i=1}^{4} \left(C_{i,1/2}\right)^{1/2} = \sqrt{m_1} + \sqrt{m_2} \leq \sqrt{2}\sqrt{m} < 2\sqrt{m}.$$

Finally, suppose $\#\mathcal{I}_\eta(m) = 1$ so that

$$\sum_{i=1}^{4} \left(C_{i,1/2}\right)^{1/2} \leq 2\sqrt{m},$$

where the equality holds if and only if $m = (0, m, 0, 0)$ or $(0, 0, m, 0)$. Therefore, these are global maximizers for the case $N = 4$. $\square$

We next consider the case $N = 3p + 1$ with $p \geq 2$. Define $m_{3p+1}^k \in \mathcal{M}_\eta$ $(k = 0, 1, \ldots, p)$ by

$$m_{3p+1}^k = (\underbrace{P_m, \ldots, P_m}_{k}, \overset{3k+1}{\check{0}}, \underbrace{P_m, \ldots, P_m}_{p-k}),$$

and $\mathcal{M}_{3p+1}^r$ by

$$\mathcal{M}_{3p+1}^r := \{m_{3p+1}^k \in \mathcal{M}_\eta : r \leq k \leq p - r\}.$$

Proposition 3 *For every $m \in \mathcal{M}_{3p+1}^0$, there exists a positive constant $\delta_{N,m,\eta} > 0$ such that $U(m, \delta) > U(\tilde{m}, \delta)$ holds for $\delta \in (0, \delta_{N,m,\eta})$ and $\tilde{m} \in \mathcal{M}_\eta \setminus \mathcal{M}_{3p+1}^0$.*

Proof The idea of this proof is the same as that of Theorem 2 (i). First, suppose $\#\mathcal{I}_\eta(m) < p$ so that

$$\sum_{i=1}^{N} \left(C_{i,1/2}\right)^{1/2} \leq \sqrt{2(2p-2)m} < 2\sqrt{p}\sqrt{m}.$$

Second, suppose $\#\mathcal{I}_\eta(m) > p$. Then $\#\mathcal{I}_0^*(m)$ is at most $3p + 1 - \#\mathcal{I}_\eta(m)$ so that

$$\begin{aligned}\sum_{i=1}^{N} \left(C_{i,1/2}\right)^{1/2} &\leq \sqrt{(3p+1-\#\mathcal{I}_\eta(m)) \sum_{i \in \mathcal{I}_0(m)} C_{i,1/2}} \\ &< \sqrt{(3p+1-\#\mathcal{I}_\eta(m))2m} \leq \sqrt{2(2p)m} = 2\sqrt{p}\sqrt{m}.\end{aligned}$$

Finally, suppose $\#\mathcal{I}_\eta(m) = p$. Then $\#\mathcal{I}_0^*(m)$ is at most $2p$ so that

$$\sum_{i=1}^{N} \left(C_{i,1/2}\right)^{1/2} \leq \sqrt{2p \sum_{i \in \mathcal{I}_0(m)} C_{i,1/2}} \leq \sqrt{2p(2m)} = 2\sqrt{p}\sqrt{m}.$$

Therefore, if the equality holds in the last case for some m, then such m must be a global maximizer.

Now we choose $m \in \mathcal{M}_\eta$ as

$$(\tilde{P}_1, \ldots, \tilde{P}_k, \overset{3k+1}{\check{0}}, \tilde{P}_{k+1}, \ldots, \tilde{P}_p),$$

where $k \in \mathbb{Z} \cup [0, p]$. Then $\#\mathcal{I}_0^*(m) = 2p$ so that

$$\begin{aligned}\sum_{i=1}^{N} \left(C_{i,1/2}\right)^{1/2} &= 2(\sqrt{m_1} + \sqrt{m_2} + \cdots + \sqrt{m_p}) \\ &\leq 2\sqrt{p(m_1 + m_2 + \cdots + m_p)} = 2\sqrt{p}\sqrt{m},\end{aligned}$$

where the equality holds if and only if $m_1 = m_2 = \cdots = m_p = m/p > 0$. Thus it is shown that any global maximizer of the total population must satisfy $m \in \mathcal{M}_{3p+1}^0$. □

We finally consider the case $N = 3p + 2$. We begin by proving the case $N = 5$.

Proof of Theorem 2 (iii-a). The idea of this proof is the same as that of Theorem 2 (i). First, suppose $\#\mathcal{I}_\eta(m) = 1$ so that

$$\sum_{i=1}^{5} \left(C_{i,1/2}\right)^{1/2} \leq 2\sqrt{m} < (1+\sqrt{2})\sqrt{m}.$$

Second, suppose $\#\mathcal{I}_\eta(m) \geq 3$ so that

$$\sum_{i=1}^{5} \left(C_{i,1/2}\right)^{1/2} \leq \sqrt{m_1+m_2} + \sqrt{m_2+m_3}$$
$$< 2\sqrt{m} < (1+\sqrt{2})\sqrt{m}, \quad (\#\mathcal{I}_\eta(m) = 3),$$

and

$$\sum_{i=1}^{5} \left(C_{i,1/2}\right)^{1/2} < \sqrt{m} < (1+\sqrt{2})\sqrt{m}, \quad (\#\mathcal{I}_\eta(m) \geq 4),$$

respectively. Finally, suppose $\#\mathcal{I}_\eta(m) = 2$ so that

$$\sum_{i=1}^{5} \left(C_{i,1/2}\right)^{1/2} \leq \sqrt{m_i} + \sqrt{m_j} + \sqrt{m_i+m_j}$$
$$\leq (1+\sqrt{2})\sqrt{m}, \quad (1 \leq i < j \leq 5),$$

where the equality holds if and only if $m = (0, m/2, 0, m/2, 0)$. Therefore, this is a unique global maximizer for the case $N = 5$. $\square$

Finally we consider the case $N = 3p + 2$ with $p \geq 2$. Define $m^k_{3p+2} \in \mathcal{M}_\eta$ $(k = 0, 1, \ldots, p-1)$ by

$$m^k_{3p+2} = (\underbrace{P_{m_*}, \ldots, P_{m_*}}_{k}, m^*, \underbrace{P_{m_*}, \ldots, P_{m_*}}_{(p-1)-k}),$$

and $\mathcal{M}^r_{3p+2}$ by

$$\mathcal{M}^r_{3p+2} := \{m^k_{3p+2} \in \mathcal{M}_\eta : r \leq k \leq (p-1) - r\},$$

where $r \in \{0, 1, \ldots, p-1\}$.

Proposition 4 *For every $m \in \mathcal{M}^0_{3p+2}$, there exists a positive constant $\delta_{N,m,\eta} > 0$ such that $U(m,\delta) > U(\tilde{m},\delta)$ holds for $\delta \in (0, \delta_{N,m,\eta})$ and $\tilde{m} \in \mathcal{M} \setminus \mathcal{M}^0_{3p+2}$.*

Proof The idea of this proof is the same as that of Theorem 2 (i). First, suppose $\#\mathcal{I}_\eta(m) < p + 1$. Then $\#\mathcal{I}^*_0(m)$ is at most $2\#\mathcal{I}_\eta$ so that

$$\sum_{i=1}^{N} \left(C_{i,1/2}\right)^{1/2} \leq \sqrt{2\#\mathcal{I}_\eta(m) \sum_{i\in\mathcal{I}_0(m)} C_{i,1/2}}$$
$$\leq \sqrt{2\#\mathcal{I}_\eta(m) 2m} \leq \sqrt{2(2p)m} = \sqrt{4pm}.$$

Second, suppose $\#\mathcal{I}_\eta(m) > p+1$. Then $\#\mathcal{I}_0^*(m)$ is at most $3p+2-\#\mathcal{I}_\eta(m)$ so that

$$\sum_{i=1}^{N}\left(C_{i,1/2}\right)^{1/2} \leq \sqrt{(3p+2-\#\mathcal{I}_\eta(m))\sum_{i\in\mathcal{I}_0(m)} C_{i,1/2}}$$
$$< \sqrt{(3p+2-\#\mathcal{I}_\eta(m))2m} \leq \sqrt{2(2p)m} = \sqrt{4pm}.$$

Finally, suppose $\#\mathcal{I}_\eta(m) = p+1$. Then $\#\mathcal{I}_0^*(m)$ is at most $2p+1$. Note that there exists at least one index $l \in \mathcal{I}_0(m)$ such that $C_{l,1/2} = m_{l-1} + m_{l+1}$.

Now we choose $m \in \mathcal{M}_\eta$ as

$$(\tilde{P}_1, \ldots, \tilde{P}_{k-1}, \tilde{m_k}, \tilde{P}_{k+2}, \ldots, \tilde{P}_{p+1}),$$

where $k \in \mathbb{Z} \cap [1, p]$, $\tilde{m}_k := (0, m_k, 0, m_{k+1}, 0)$. Then $\#\mathcal{I}_0^*(m) = 2p+1$ so that

$$\sum_{i=1}^{N}\left(C_{i,1/2}\right)^{1/2} = \sqrt{m_k} + \sqrt{m_{k+1}} + \sqrt{m_k + m_{k+1}} + \sum_{i\in\{1,2,\ldots,p+1\}\setminus\{k,k+1\}} 2\sqrt{m_i}.$$

Define the function $f^k, g : \mathbb{R}^{p+1} \to \mathbb{R}$ as

$$f^k := \sqrt{x_k} + \sqrt{x_{k+1}} + \sqrt{x_k + x_{k+1}} + \sum_{i\in\{1,2,\ldots,p+1\}\setminus\{k,k+1\}} 2\sqrt{x_i},$$

and

$$g := \sum_{i=1}^{p+1} x_i - m,$$

respectively. We calculate the interior critical point of f^k under the constraint of g by using the Lagrange multipliers. then we have unique solution as follows:

$$\begin{cases} x_i = \dfrac{4m}{4(p-1)+(1+\sqrt{2})^2}, & (i \in \{1,2,\ldots,p+1\} \setminus \{k,k+1\}), \\ x_k = x_{k+1} = \dfrac{(1+\sqrt{2})^2 m}{2\{4(p-1)+(1+\sqrt{2})^2\}}. \end{cases} \tag{6.6}$$

In fact, this critical point is maximum point of f^k since Hesse matrix of f^k is negative definite. Hence (6.6) is a unique maximizer of f^k subject to $g = 0$. Thus we conclude

$$\max_{m\in\mathcal{M}_\eta} \sum_{i=1}^{N}\left(C_{i,1/2}\right)^{1/2} = \sqrt{(4p+2\sqrt{2}+1)m} > \sqrt{4pm}.$$

Thus it is shown that any global maximizer of the total population must satisfy $m \in \mathcal{M}^0_{3p+2}$. □

From Propositions 3 and 4, the calculation of the total population will be naturally divided into two part. Lemma 3 deals with the case $N = 3p + 1$, and Lemma 4 deals with the case $N = 3p + 2$. By using the definition of $\mathcal{M}_\eta$, if we choose two positive parameter η_1, η_2 satisfying $\eta_1 > \eta_2$, then $\mathcal{M}_{\eta_1} \subset \mathcal{M}_{\eta_2}$ must be hold. Further, we have $\mathcal{M}^0_{3p+1} \subset \mathcal{M}_{m/p}$ and $\mathcal{M}^0_{3p+2} \subset \mathcal{M}_{m^*}$. Hence any global maximizer of the total population within $\mathcal{M}_{p^*}$ are also one within $\mathcal{M}_\eta$.

Lemma 3 *For every $m \in \mathcal{M}^1_{3p+1}$, there exists a positive constant $\delta_{N,m,\eta} > 0$ such that $U(m, \delta) > U(\tilde{m}, \delta)$ holds for $\delta \in (0, \delta_{N,m,\eta})$ and $\tilde{m} \in \mathcal{M}^0_{3p+1} \setminus \mathcal{M}^1_{3p+1}$.*

Proof We calculate the total population $U(\tilde{m}, \delta)$ and $U(m, \delta)$. We use (6.5) to have

$$U(\tilde{m}, \delta) = m + 2\sqrt{p}\sqrt{m}\delta^{1/2} + \left(\frac{m}{p}\right)^{1/4} \delta^{3/4} - (3p + 1)\delta + o(\delta)$$

and

$$U(m, \delta) = m + 2\sqrt{p}\sqrt{m}\delta^{1/2} + \sqrt{2}\left(\frac{m}{p}\right)^{1/4} \delta^{3/4} - (3p + 1)\delta + o(\delta),$$

respectively. Therefore, we obtain

$$U(m, \delta) - U(\tilde{m}, \delta) = (\sqrt{2} - 1)(m/p)^{1/4}\delta^{3/4} + o(\delta) > 0.$$

This completes the proof. □

Lemma 4 *For every $m \in \mathcal{M}^1_{3p+2}$, there exists a positive constant $\delta_{N,m,\eta} > 0$ such that $U(m, \delta) > U(\tilde{m}, \delta)$ holds for $\delta \in (0, \delta_{N,m,\eta})$ and $\tilde{m} \in \mathcal{M}^0_{3p+2} \setminus \mathcal{M}^1_{3p+2}$.*

This can be proved in the same manner as Lemma 3, so we omit the proof.

7 Proof of Theorem 2

7.1 Computation of higher order terms

In this subsection, we compute a coefficient of $\delta^{n/4}$, where $n \in \mathbb{Z}_{\geq 1}$. To compare the total population when the resource allocation pattern is included in $\mathcal{M}^0_{3p+1}$ or $\mathcal{M}^0_{3p+2}$, we calculate an asymptotic expansion with respect to δ. If the resources are placed at intervals of 4 patches or less, Proposition 7 allows asymptotic expansion in the order of $\delta^{n/4}$. Lemmas 3 and 4 indicate that a global maximizer of $U(m, \delta)$ must belong to $\mathcal{M}^1_{3p+1}$ or $\mathcal{M}^1_{3p+2}$. This means that the resources of the maximizer are placed at intervals of 4 patches or less. At the end of this section, we identify the resource allocation that maximizes the total population by comparing the coefficients on the order of $\delta^{n/4}$. In view of (6.4), we already have expansion of the positive solution of

(1.3) in the case $m \in \mathcal{M}^1_{3p+1}$ or $\mathcal{M}^1_{3p+2}$ as

$$u_i = \sum_{q=0}^{4} u_{i,q}\delta^{q/4} + o(\delta).$$

Here, $u_{i,q}$ is a coefficient of u_i of $\delta^{q/4}$ depending on m and N.

In the next lemma, we expand u_i up to a higher order of δ.

Lemma 5 *Let n be a positive integer and choose $m \in \mathcal{M}^1_{3p+1}$ or $m \in \mathcal{M}^1_{3p+2}$ arbitrarily. Then the solution of (1.3) is expanded as*

$$u_i = \sum_{q=0}^{n} u_{i,q}\delta^{q/4} + o(\delta^{n/4})$$

for all $i \in \Omega$.

Proof We proceed by induction on n. Suppose that

$$u_i = \sum_{q=0}^{n} u_{i,q}\delta^{q/4} + o(\delta^{n/4})$$

for all $i \in \Omega$ and $n \geq 5$. We divide the proof into three parts according to the value of $\underline{\deg}\, u_i$.

First, we choose $i \in \Omega$ such that $\underline{\deg}\, u_i = 0$. By (3.2), we have

$$u_i = \frac{m_i - 2\delta}{2} + \left\{ \frac{(m_i - 2\delta)^2}{4} + \delta\left(\sum_{q=0}^{n}(u_{i-1,q} + u_{i+1,q})\delta^{q/4} + o(\delta^{n/4})\right)\right\}^{1/2}.$$

By our assumption, we have $\underline{\deg}\, u_{i-1} = \underline{\deg}\, u_{i+1} = 1/2$ to compute

$$\begin{aligned} u_i &= \frac{m_i - 2\delta}{2} + \left\{ \frac{(m_i - 2\delta)^2}{4} + \delta\left(\sum_{q=2}^{n}(u_{i-1,q} + u_{i+1,q})\delta^{q/4} + o(\delta^{n/4})\right)\right\}^{1/2} \\ &= \frac{m_i - 2\delta}{2} + \left\{ \frac{(m_i - 2\delta)^2}{4} + \sum_{q=6}^{n+4}(u_{i-1,q-4} + u_{i+1,q-4})\delta^{q/4} + o(\delta^{(n+4)/4})\right\}^{1/2}. \end{aligned}$$

Using binomial series and uniqueness of u_i, we obtain

$$u_i = \sum_{q=0}^{n+4} u_{i,q}\delta^{q/4} + o(\delta^{(n+4)/4}).$$

Second, we choose $i \in \Omega$ such that $\underline{\deg}\, u_i = 1/2$. From our assumption, we see that this condition is satisfied for $i = 1$ and $i = N$. By (3.2), we have

$$u_i = -\delta + \left\{\delta^2 + \delta\left(\sum_{q=0}^{n}(u_{i-1,q} + u_{i+1,q})\delta^{q/4} + o(\delta^{n/4})\right)\right\}^{1/2}.$$

We use binomial series and uniqueness of u_i to obtain

$$u_i = \delta^{1/2}\left(\sum_{q=0}^{n} u_{i,q+2}\delta^{q/4} + o(\delta^{n/4})\right) = \sum_{q=0}^{n+2} u_{i,q}\delta^{q/4} + o(\delta^{(n+2)/4}).$$

Calculation of u_1 and u_N is the same as above, so we omit the proof.

Finally, we choose $i \in \Omega$ such that $\underline{\deg}\, u_i = 3/4$. By (3.2), we have

$$u_i = -\delta + \left\{\delta^2 + \delta\left(\sum_{q=0}^{n}(u_{i-1,q} + u_{i+1,q})\delta^{q/4} + o(\delta^{n/4})\right)\right\}^{1/2}.$$

By our assumption, we have $\underline{\deg}\, u_{i-1} = \underline{\deg}\, u_{i+1} = 1/2$. Using binomial series and uniqueness of u_i, we obtain

$$\begin{aligned} u_i &= -\delta + \left\{\delta^{3/2}\left(\delta^{1/2} + \sum_{q=0}^{n-2}(u_{i-1,q+2} + u_{i+1,q+2})\delta^{q/4} + o(\delta^{(n-2)/4})\right)\right\}^{1/2} \\ &= \delta^{3/4}\left(\sum_{q=0}^{n-2} u_{i,q+3}\delta^{q/4} + o(\delta^{(n-2)/4})\right) = \sum_{q=0}^{n+1} u_{i,q}\delta^{q/4} + o(\delta^{(n+1)/4}). \end{aligned}$$

This completes the proof. □

From the second equation in (1.3), we may assume $u_{0,q} = u_{1,q}$ and $u_{N+1,q} = u_{N,q}$. In view of Lemma 5, the positive solution of (1.3) must satisfy

$$\begin{aligned} &m_i\left(\sum_{q=0}^{n} u_{i,q}\delta^{q/4} + o(\delta^{n/4})\right) - \left(\sum_{q=0}^{n} u_{i,q}\delta^{q/4} + o(\delta^{n/4})\right)^2 \\ &\quad + \delta\left(\sum_{q=0}^{n}(u_{i-1,q} + u_{i+1,q} - 2u_{i,q})\delta^{q/4} + o(\delta^{n/4})\right) = 0 \end{aligned}$$

for all $i \in \Omega$ and $n \in \mathbb{Z}_+$. Rearranging this equation in ascending powers of δ up to the term in $\delta^{n/4}$, we obtain

$$\sum_{q=0}^{3}\left(m_i u_{i,q} - \sum_{j=0}^{q} u_{i,j}u_{i,q-j}\right)\delta^{q/4}$$
$$+\sum_{q=4}^{n}\left(m_i u_{i,q} - \left(\sum_{j=0}^{q} u_{i,j}u_{i,q-j}\right) + (u_{i-1,q-4} + u_{i+1,q-4} - 2u_{i,q-4})\right)\delta^{q/4}$$
$$+ o(\delta^{n/4}) = 0. \tag{7.1}$$

Lemma 6 *Let $n \geq 5$ and q be a positive integer. Choose $m \in \mathcal{M}^1_{3p+1}$, or $m \in \mathcal{M}^1_{3p+2}$ arbitrarily. Then coefficients of the expansion of the positive solution of* (1.3) *satisfy the following equalities:*

$$\begin{cases} m_i u_{i,q} - \sum_{j=0}^{q} u_{i,j}u_{i,q-j} = 0, \quad (q = 0, 1, 2, 3), \\ m_i u_{i,q} - \left(\sum_{j=0}^{q} u_{i,j}u_{i,q-j}\right) + (u_{i-1,q-4} + u_{i+1,q-4} - 2u_{i,q-4}) = 0, \\ \quad (q = 4, 5, \ldots, n). \end{cases} \tag{7.2}$$

Proof To prove the cases $q = 0, \ldots, 4$, we substitute every initial term of $u_{i,q}$ to (7.2) directly, which was determined in Sect. 6. To prove the case $q = 5, 6, \ldots, n$, we use induction on q. Suppose that (7.2) holds $q = 1, 2, \ldots, s-1$, where s is a positive integer satisfying $s \leq n$. Substituting (7.2) to (7.1), we have

$$\sum_{q=s}^{n}\left(m_i u_{i,q} - \left(\sum_{j=0}^{q} u_{i,j}u_{i,q-j}\right) + (u_{i-1,q-4} + u_{i+1,q-4} - 2u_{i,q-4})\right)\delta^{q/4}$$
$$+ o(\delta^{n/4}) = 0.$$

Dividing this by $\delta^{s/4}$, we have

$$\sum_{q=s}^{n}\left(m_i u_{i,q} - \left(\sum_{j=0}^{q} u_{i,j}u_{i,q-j}\right) + (u_{i-1,q-4} + u_{i+1,q-4} - 2u_{i,q-4})\right)\delta^{(q-s)/4}$$
$$+ o(\delta^{(n-s)/4}) = 0.$$

Since this equality must hold for any $\delta > 0$ small, we conclude the proof. □

Hereafter, we assume that n is sufficiently large. We prepare some recurrence relation about $u_{i,q}$ for every $i \in \Omega$ and $q = 5, 6, \ldots, n$, which will be used in Sects. 7.2 and 7.3.

We choose $i \in \Omega$ such that $\underline{\deg}\, u_i = 0$. In this case, the second equality of (7.2) becomes

$$m_i u_{i,q} - \left(2u_{i,0}u_{i,q} + \sum_{j=1}^{q-1} u_{i,j}u_{i,q-j} \right) + (u_{i-1,q-4} + u_{i+1,q-4} - 2u_{i,q-4}) = 0.$$

By (6.4), we have $u_{i,0} = m_i$. Therefore, $u_{i,q}$ can be expressed by recurrence relation as

$$u_{i,q} = {m_i}^{-1} \left\{ - \left(\sum_{j=1}^{q-1} u_{i,j}u_{i,q-j} \right) + (u_{i-1,q-4} + u_{i+1,q-4} - 2u_{i,q-4}) \right\}. \tag{7.3}$$

We can show the recurrence relation of $u_{i,q}$ satisfying $\underline{\deg}\, u_i = 1/2$ and $\underline{\deg}\, u_i = 3/4$ as

$$u_{i,q} = (2u_{i,2})^{-1} \left\{ - \left(\sum_{j=3}^{q-1} u_{i,j}u_{i,q+2-j} \right) + (u_{i-1,q-2} + u_{i+1,q-2} - 2u_{i,q-2}) \right\} \tag{7.4}$$

and

$$u_{i,q} = (2u_{i,3})^{-1} \left\{ - \left(\sum_{j=4}^{q-1} u_{i,j}u_{i,q+3-j} \right) + (u_{i-1,q-1} + u_{i+1,q-1} - 2u_{i,q-1}) \right\}, \tag{7.5}$$

respectively. Since the proofs of (7.4) and (7.5) can be obtained in the same way as that of (7.3), we omit the proof.

7.2 Proof of Theorem 2 (ii-b) and (ii-c)

Our objective in this subsection is to express $u_{i,q}$ by using some recurrence relation, and show that $U(m^{r+1}_{3p+1}, \delta) - U(m^r_{3p+1}, \delta)$ is positive for some r.

Let k and $r \in [1, \lfloor p/2 \rfloor]$ be positive integers, where $\lfloor \cdot \rfloor$ is a floor function. We choose $m^r_{3p+1} \in \mathcal{M}^1_{3p+1}$ and define

$$\begin{cases} i^r_+(k) := 3r + 1 + k, & (1 \le k \le 3(p - r)), \\ i^r(0) := 3r + 1, \\ i^r_-(k) := 3r + 1 - k, & (1 \le k \le 3r). \end{cases}$$

Further, we define $i^r_\pm(k) := i^r_-(k)$ or $i^r_+(k)$. Let $\{a_k\}$ and $\{b_k\}$ be monotone increasing sequences given by

$$\begin{cases} a_k := 3k-1, \\ b_k := \min_{2\le k} \left\{ (\mathbb{Z}_{\ge 1} \setminus \{a_\lambda\}_{\lambda\ge 1}) \setminus \bigcup_{\lambda=1}^{k-1} b_\lambda \right\}, b_1 = 1. \end{cases}$$

We define the sets

$$\mathbb{A}_r := \{a_k \in \mathbb{Z} \mid 1 \le k \le r\}, \quad \mathbb{B}_r := \{b_k \in \mathbb{Z} \mid 1 \le k \le r\}.$$

We also define $\{A_q\}_{q\in\mathbb{Z}_{\ge 0}}$ and $\{B_q\}_{q\in\mathbb{Z}_{\ge 0}}$ by the recurrence relation

$$\begin{aligned} A_q &= \left(\frac{m}{p}\right)^{-1} \left\{ -\sum_{j=1}^{q-1} A_j A_{q-j} + (B_{q-4} + B_{q-4} - 2A_{q-4}) \right\} \text{ for } q \ge 5, \\ B_q &= \frac{1}{2}\left(\frac{m}{p}\right)^{-1/2} \left\{ -\sum_{j=3}^{q-1} B_j B_{q+2-j} + (A_{q-2} + B_{q-2} - 2B_{q-2}) \right\} \text{ for } q \ge 5, \end{aligned}$$

and the initial conditions

$$\begin{aligned} &A_0 = \frac{m}{p}, \quad A_1 = A_2 = A_3 = 0, \quad A_4 = -2, \\ &B_0 = B_1 = 0, \quad B_2 = (m/p)^{1/2}, \quad B_3 = 0, \quad B_4 = -\frac{1}{2}. \end{aligned}$$

Define $\{C_{0,q}\}_{q\in\mathbb{Z}_{\ge 0}}$ by the recurrence relation

$$\begin{aligned} C_{0,q} = \frac{1}{2\sqrt{2}}\left(\frac{m}{p}\right)^{-1/4} \Bigg\{ &- \sum_{j=4}^{q-1} (C_{0,j} C_{0,q+3-j}) \\ &+ (B_{q-1} + E_{1,q-1}) + (B_{q-1} + E_{1,q-1}) - 2C_{0,q-1} \Bigg\} \text{ for } q \ge 5, \end{aligned}$$

and the initial condition

$$C_{0,0} = C_{0,1} = C_{0,2} = 0, \quad C_{0,3} = \sqrt{2}\left(\frac{m}{p}\right)^{1/4}, \quad C_{0,4} = -1.$$

Define $\{E_{k,q}\}_{(k,q)\in\mathbb{Z}_{\geq 1}\times\mathbb{Z}_{\geq 0}}$ by the recurrence relation

$$\begin{aligned}
E_{1,q} &= \frac{1}{2}\left(\frac{m}{p}\right)^{-1/2}\Bigg\{-\sum_{j=3}^{q-1}(B_jE_{1,q+2-j}+E_{1,j}B_{q+2-j}+E_{1,j}E_{1,q+2-j})\\
&\quad+(C_{0,q-2}-B_{q-2})+E_{2,q-2}-2E_{1,q-2}\Bigg\}\text{ for } q\geq 5,\\
E_{a_k,q} &= \left(\frac{m}{p}\right)^{-1}\Bigg\{-\sum_{j=1}^{q-1}(A_jE_{a_k,q-j}+E_{a_k,j}A_{q-j}+E_{a_k,j}E_{a_k,q-j})\\
&\quad+(E_{a_k-1,q-4}+E_{a_k+1,q-4}-2E_{a_k,q-4})\Bigg\}\text{ for } q\geq 5, k\geq 1,\\
E_{b_k,q} &= \frac{1}{2}\left(\frac{m}{p}\right)^{-1/2}\Bigg\{-\sum_{j=3}^{q-1}(B_jE_{b_k,q+2-j}+E_{b_k,j}B_{q+2-j}+E_{b_k,j}E_{b_k,q+2-j})\\
&\quad+(E_{b_k-1,q-2}+E_{b_k+1,q-2}-2E_{b_k,q-2})\Bigg\}\text{ for } q\geq 5, k\geq 1,
\end{aligned}$$

and the initial condition

$$\begin{aligned}
&E_{1,0}=E_{1,1}=E_{1,2}=E_{1,3}=0,\quad E_{1,4}=-\frac{1}{2},\\
&E_{k,0}=E_{k,1}=E_{k,2}=E_{k,3}=E_{k,4}=0\quad\text{for } k\geq 2.
\end{aligned}$$

Finally, we define a mapping $k:\mathbb{Z}_{\geq 2}\to\mathbb{Z}_{\geq 0}$ by

$$k(\eta)=\begin{cases}3d-1 & \text{if } \eta=8d, 8d+1,\\ 3d & \text{if } \eta=8d+2, 8d+3,\\ 3d+1 & \text{if } \eta=8d+4, 8d+5, 8d+6, 8d+7,\end{cases}$$

and its inverse by

$$d(k):=\min k^{-1}(\{k\}).$$

From the definition of $E_{k,q}$, we have

$$E_{k,q}=0\text{ for } 0\leq q\leq d(k)-1, \tag{7.6}$$

$$E_{a_k,d(a_k)}=\left(\frac{m}{p}\right)^{-1}E_{a_k-1,d(a_k-1)}, \tag{7.7}$$

$$E_{b_k,d(b_k)}=\frac{1}{2}\left(\frac{m}{p}\right)^{-1/2}E_{b_k-1,d(b_k-1)}\quad\text{for } k\geq 2. \tag{7.8}$$

We first prove that $u_{i,q}$ can be expressed by using A_q, B_q, and $E_{k,q}$.

Proposition 5 *Let q be an integer and choose $m \in \mathcal{M}^1_{3p+1}$ arbitrarily. If $0 \leq q \leq d(6r)+1$, the coefficient of the expanded positive solution of (1.3) can be expressed as*

$$u_{i^r_-(a_k),q} = A_q + E_{a_k,q} + E_{6r+1-a_k,q} \quad \textit{for } a_k \in \mathbb{A}_r, \tag{7.9}$$

$$u_{i^r_-(b_k),q} = B_q + E_{b_k,q} + E_{6r+1-b_k,q} \quad \textit{for } b_k \in \mathbb{B}_{2r}, \tag{7.10}$$

$$u_{i^r_+(a_k),q} = A_q + E_{a_k,q} + E_{6(p-r)+1-a_k,q} \quad \textit{for } a_k \in \mathbb{A}_{p-r}, \tag{7.11}$$

$$u_{i^r_+(b_k),q} = B_q + E_{b_k,q} + E_{6(p-r)+1-b_k,q} \quad \textit{for } b_k \in \mathbb{B}_{2(p-r)}. \tag{7.12}$$

Moreover,

$$u_{i^r(0),q} = \begin{cases} C_{0,q} & \textit{for } 0 \leq q \leq d(6r), \\ C_{0,q} + E_0 & \textit{for } q = d(6r)+1, \end{cases} \tag{7.13}$$

where $E_0 := (2\sqrt{2})^{-1}(m/p)^{-1/4}(E_{6r,d(6r)} + E_{6(p-r),d(6r)}) < 0$ is a constant.

Proof We first show (7.9). To prove the case $q = 0, 1, \ldots, 4$, we substitute every initial term of $u_{i,q}$ to (7.9)–(7.13) directly, which was determined in Sect. 6. To prove the case $q = 5, 6, \ldots, d(6r)+1$, we use induction on q. Suppose that (7.9)–(7.13) hold for $q = 1, 2, \ldots, s-1$, where s is positive integer satisfying $s \leq d(6r)+1$. Then the first term of the right-hand side of (7.3) becomes

$$\begin{aligned} &\sum_{j=1}^{s-1} u_{i^r_-(a_k),j} u_{i^r_-(a_k),s-j} \\ &\quad = -\sum_{j=1}^{s-1} (A_j A_{s-j} + A_j E_{a_k,s-j} + A_j E_{6r+1-a_k,s-j} \\ &\qquad + E_{a_k,j} A_{s-j} + E_{a_k,j} E_{a_k,s-j} + E_{a_k,j} E_{6r+1-a_k,s-j} \\ &\qquad + E_{6r+1-a_k,j} A_{s-j} + E_{6r+1-a_k,j} E_{a_k,s-j} + E_{6r+1-a_k,j} E_{6r+1-a_k,s-j}). \end{aligned} \tag{7.14}$$

We claim that

$$E_{a_k,j} E_{6r+1-a_k,s-j} = E_{6r+1-a_k,j} E_{a_k,s-j} = 0 \tag{7.15}$$

for all $1 \leq j \leq s-1$. The proof of this claim will be divided into three cases according to s. First, suppose $5 \leq s \leq d(a_k)$. In this case, j satisfies $1 \leq j \leq d(a_k)-1$. This gives $E_{a_k,j} = E_{6r+1-a_k,j} = 0$ by (7.6). Second, suppose $d(a_k)+1 \leq s \leq d(6r+1-a_k)$. If $1 \leq j \leq d(a_k)-1$, then we can prove (7.15) in the same manner as the first case. If $d(a_k) \leq j \leq s-1$, then $d(6r+1-a_k)-1$ is the upper bound of j. This gives $E_{6r+1-a_k,j} = 0$ by (7.6). Moreover, $d(6r+1-a_k)-1$ is the upper bound of $s-j$, since

$$\begin{aligned} s - j &\leq s - d(a_k) \leq d(6r+1-a_k) - d(a_k) = d(6r-2a_k) \\ &< d(6r+1-a_k) - 1. \end{aligned}$$

From (7.6), we have $E_{6r+1-a_k,q-j}=0$. Finally, suppose $d(6r+1-a_k)+1\le s\le d(6r)+1$. If $1\le j\le d(6r+1-a_k)-1$, then we use the same manner as the second case. If $d(6r+1-a_k)\le j\le s-1$, then $d(a_k)-1$ is the upper bound of $s-j$, since

$$\begin{aligned}s-j&\le s-d(6r+1-a_k)\le d(6r)+1-d(6r+1-a_k)\\&=8(k-1)+3=d(a_k-2)+1<d(a_k)-1.\end{aligned}$$

In view of (7.6), $E_{a_k,q-j}=E_{6r+1-a_k,q-j}=0$. Hence we obtain (7.15).

Thus, (7.14) can be rewritten as

$$\begin{aligned}&\sum_{j=1}^{s-1}u_{i_-^r(a_k),j}u_{i_-^r(a_k),s-j}\\&\quad=-\sum_{j=1}^{s-1}(A_jA_{s-j})-\sum_{j=1}^{s-1}(A_jE_{a_k,s-j}+E_{a_k,j}A_{s-j}+E_{a_k,j}E_{a_k,s-j})\\&\qquad-\sum_{j=1}^{s-1}(A_jE_{6r+1-a_k,s-j}+E_{6r+1-a_k,j}A_{s-j}+E_{6r+1-a_k,j}E_{6r+1-a_k,s-j}).\end{aligned}\tag{7.16}$$

The rest of the proof of (7.9) is to calculate the second term of the right-hand side of (7.3). We have

$$\begin{aligned}&u_{i_-^r(a_k)-1,s-4}+u_{i_-^r(a_k)+1,s-4}-2u_{i_-^r(a_k),s-4}\\&\quad=(B_{s-4}+B_{s-4}-2A_{s-4})+(E_{a_k-1,s-4}+E_{a_k+1,s-4}-2E_{a_k,s-4})\\&\qquad+(E_{6r+1-a_k-1,s-4}+E_{6r+1-a_k+1,s-4}-2E_{6r+1-a_k,s-4}).\end{aligned}\tag{7.17}$$

Combining (7.16) and (7.17), we obtain (7.9).

Next, let us show (7.10) in the same way as the proof of (7.9). Suppose that (7.9)–(7.13) hold for $q=1,2,\ldots,s-1$. Then the first term of the right-hand side of (7.4) becomes

$$\begin{aligned}&\sum_{j=3}^{s-1}u_{i_-^r(b_k),j}u_{i_-^r(b_k),s+2-j}\\&\quad=-\sum_{j=3}^{s-1}(B_jB_{s+2-j}+B_jE_{b_k,s+2-j}+B_jE_{6r+1-b_k,s+2-j}\\&\qquad+E_{b_k,j}B_{s+2-j}+E_{b_k,j}E_{b_k,s+2-j}+E_{b_k,j}E_{6r+1-b_k,s+2-j}\\&\qquad+E_{6r+1-b_k,j}B_{s+2-j}+E_{6r+1-b_k,j}E_{b_k,s+2-j}+E_{6r+1-b_k,j}E_{6r+1-b_k,s+2-j}).\end{aligned}$$

We also claim that

$$E_{b_k,j}E_{6r+1-b_k,s+2-j}=E_{6r+1-b_k,j}E_{b_k,s+2-j}=0\tag{7.18}$$

for all $3 \le j \le s-1$.

The proof of this claim will be also divided into three cases depending to s. First, suppose $5 \le s \le d(b_k)$. In this case, j satisfies $3 \le j \le d(b_k)-1$. This gives $E_{b_k,j} = E_{6r+1-b_k,j} = 0$ by (7.6). Second, suppose $d(b_k)+1 \le s \le d(6r+1-b_k)$. If $3 \le j \le d(b_k)-1$, then we use the same manner as the first case. If $d(b_k) \le j \le s-1$, then $d(6r+1-b_k)-1$ is the upper bound of j. This gives $E_{6r+1-b_k,j} = 0$ by (7.6). Moreover, $d(6r+1-b_k)-1$ is an upper bound of $s+2-j$, since

$$\begin{aligned} s+2-j \le s+2-d(b_k) &\le d(6r+1-b_k)+2-d(b_k) \\ &= d(6r+1-2b_k) < d(6r+1-b_k)-1. \end{aligned}$$

From (7.6), we have $E_{6r+1-b_k,s-j} = 0$. Finally, suppose $d(6r+1-b_k)+1 \le s \le d(6r)+1$. If $1 \le j \le d(6r+1-b_k)-1$, then we use the same manner as the second case. If $d(6r+1-b_k) \le j \le s-1$, then $d(b_k)-1$ is an upper bound of $s+2-j$, since

$$\begin{aligned} s+2-j \le s+2-d(6r+1-b_k) &\le d(6r)+3-d(6r+1-b_k) \\ &= d(b_k-1)+1 \le d(b_k)-1. \end{aligned}$$

In view of (7.6), $E_{b_k,s+2-j} = E_{6r+1-b_k,s+2-j} = 0$. Hence we obtain (7.18). Thus, (7.14) can be rewritten as

$$\begin{aligned} &\sum_{j=3}^{s-1} u_{i_-^r(b_k),j} u_{i_-^r(b_k),s+2-j} \\ &\quad = -\sum_{j=3}^{s-1}(B_j B_{s+2-j}) - \sum_{j=3}^{s-1}(B_j E_{b_k,s+2-j} + E_{b_k,j} B_{s+2-j} + E_{b_k,j} E_{b_k,s+2-j}) \\ &\qquad - \sum_{j=3}^{s-1}(B_j E_{6r+1-b_k,s+2-j} + E_{6r+1-b_k,j} B_{s-j} + E_{6r+1-b_k,j} E_{6r+1-b_k,s+2-j}). \end{aligned} \tag{7.19}$$

The rest of the proof of (7.10) is to calculate the second term of the right-hand side of (7.4). We have

$$\begin{aligned} &u_{i_-^r(b_k)-1,s-2} + u_{i_-^r(b_k)+1,s-2} - 2u_{i_-^r(b_k),s-2} \\ &\quad = (A_{s-2} + B_{s-2} - 2B_{s-2}) + (E_{b_k-1,s-2} + E_{b_k+1,s-2} - 2E_{b_k,s-2}) \\ &\qquad + (E_{6r+1-b_k-1,s-2} + E_{6r+1-b_k+1,s-2} - 2E_{6r+1-b_k,s-2}). \end{aligned} \tag{7.20}$$

Combining (7.19) and (7.20), we obtain (7.10). We can apply the same manner as above to obtain representation formula (7.11) and (7.12). So we omit the proof.

Finally, we prove (7.13) by induction on q. Suppose that (7.9)–(7.13) hold for $q = 1, 2, \dots, s-1$. By (7.5), we have

$$u_{i^r(0),s} = \frac{1}{2\sqrt{2}}\left(\frac{m}{p}\right)^{-1/4}\Bigg\{-\left(\sum_{j=4}^{s-1} C_{0,j}C_{0,s+3-j}\right)$$

$$+ (B_{s-1} + E_{1,s-1} + E_{6r,s-1}) + (B_{s-1} + E_{1,s-1} + E_{6(p-r),s-1}) - 2C_{0,s-1})\Bigg\}$$

$$= C_{0,s} + \frac{1}{2\sqrt{2}}\left(\frac{m}{p}\right)^{-1/4}(E_{6r,s-1} + E_{6(p-r),s-1}).$$

In view of (7.6), $E_{6r,s-1} = E_{6(p-r),s-1} = 0$ for $4 \le s-1 \le d(6r)-1$. Further, we have $E_{6r,d(6r)} < 0$ by using (7.7)–(7.8) and the initial condition $E_{1,4}$, so that

$$u_{i^r(0),q} = C_{0,q} + \frac{1}{2\sqrt{2}}\left(\frac{m}{p}\right)^{-1/4}(E_{6r,d(6r)} + E_{6(p-r),d(6r)}) \quad \text{if } q = d(6r)+1.$$

Thus, we obtain (7.13) and conclude the proof. □

Let us complete the proof of Theorem 2 (ii). We consider $U(m^r_{3p+1}, \delta)$ for $1 \le r < p/2 - 1$. By Proposition 5 and (7.6), we have

$$U(m^r_{3p+1}, \delta) = \sum_{q=0}^{d(6r)+1}\sum_{i=1}^{N} u_{i,q}\delta^{q/4} + o(\delta^{(d(6r)+1)/4})$$

$$= \sum_{q=0}^{3}(C_{0,q} + pA_q + 2pB_q)\delta^{q/4}$$

$$+ \sum_{q=4}^{d(6r)}(C_{0,q} + pA_q + 2pB_q + \sum_{k=1}^{k(q)} 2E_{k,q})\delta^{q/4}$$

$$+ \sum_{q=d(6r)+1}(C_{0,q} + E_0 + pA_q + 2pB_q + \sum_{k=1}^{6r} 2E_{k,q})\delta^{q/4} + o(\delta^{(d(6r)+1)/4}).$$

Hence we obtain

$$U(m^{r+1}_{3p+1}, \delta) - U(m^r_{3p+1}, \delta) = -E_0\delta^{(d(6r)+1)/4} + o(\delta^{(d(6r)+1)/4}) > 0$$

for sufficiently small δ. This proves Theorem 2 for $N = 3p+1$. □

7.3 Proof of Theorem 2 (iii-b) and (iii-c)

This subsection is organized in the same way as Sect. 7.2. For simplicity, we use the same notation as in Sect. 7.2. Let k and $r \in [1, \lfloor (p-1)/2 \rfloor]$ be positive integer. We choose $m^r_{3p+2} \in \mathcal{M}^1_{3p+2}$ and define

$$\begin{cases} i^r_+(k) := 3r + 3 + k \quad (1 \le k \le 3\{(p-1) - r\} + 2), \\ i^r(0) := 3r + 3, \\ i^r_-(k) := 3r + 3 - k \quad (1 \le k \le 3r + 2). \end{cases}$$

We define monotone increasing sequences by

$$\begin{cases} a_k := 3k + 1 \ (k \in \mathbb{Z}_{\ge 0}), \\ b_k := \min\limits_{1 \le k} \left\{ (\mathbb{Z}_{\ge 2} \setminus \{a_\lambda\}_{\lambda \ge 1}) \setminus \bigcup\limits_{\lambda=0}^{k-1} b_\lambda \right\}, \ b_0 = 2. \end{cases}$$

We define the sets

$$\mathbb{A}_r := \{a_k \in \mathbb{Z} \mid 1 \le k \le r\}, \quad \mathbb{B}_r := \{b_k \in \mathbb{Z} \mid 1 \le k \le r\}.$$

We also define $\{A_q\}_{q \in \mathbb{Z}_{\ge 0}}$ and $\{B_q\}_{q \in \mathbb{Z}_{\ge 0}}$ by the recurrence relation

$$A_q = (m_*)^{-1} \left\{ -\sum_{j=1}^{q-1} A_j A_{q-j} + (B_{q-4} + B_{q-4} - 2A_{q-4}) \right\} \text{ for } q \ge 5,$$
$$B_q = \frac{1}{2} (m_*)^{-1/2} \left\{ -\sum_{j=3}^{q-1} B_j B_{q+2-j} + (A_{q-2} + B_{q-2} - 2B_{q-2}) \right\} \text{ for } q \ge 5,$$

and the initial condition

$$A_0 = m_*, \quad A_1 = A_2 = A_3 = 0, \quad A_4 = -2,$$
$$B_0 = B_1 = 0, \quad B_2 = (m_*)^{1/2}, \quad B_3 = 0, \quad B_4 = -\frac{1}{2}.$$

Define $\{C_{k,q}\}_{(k,q) \in \mathbb{Z} \cap [0,2] \times \mathbb{Z}_{\ge 0}}$ by the recurrence relations

$$C_{0,q} = \frac{1}{2} \left(m^*\right)^{-1/2} \Big\{ -\sum_{j=3}^{q-1} (C_{0,j} C_{0,q+2-j}) + (C_{1,q-2} + C_{1,q-2} - 2C_{0,q-2}) \Big\},$$

$$
\begin{aligned}
C_{1,q} &= \left(m^*\right)^{-1/2}\Big\{-\sum_{j=1}^{q-1}(C_{1,j}C_{1,q-j}) + (C_{0,q-2} + C_{2,q-2} - 2C_{1,q-2})\Big\},\\
C_{2,q} &= \frac{1}{2}\left(m^*\right)^{-1/2}\Big\{-\sum_{j=3}^{q-1}(C_{2,j}C_{2,q+2-j})\\
&\quad + (C_{1,q-2} + (B_{3,q-2} + E_{3,q-2}) - 2C_{2,q-2}\Big\},
\end{aligned}
$$

for $q \geq 5$, and the initial conditions

$$
\begin{aligned}
&C_{0,0} = C_{0,1} = 0, \quad C_{0,2} = (2m^*)^{1/2}, \quad C_{0,3} = 0, \quad C_{0,4} = -1,\\
&C_{1,0} = m^*, \quad C_{1,1} = C_{1,2} = C_{1,3} = 0, \quad C_{1,4} = -2,\\
&C_{2,0} = C_{2,1} = 0, \quad C_{2,2} = (m_*)^{1/2}, \quad C_{2,3} = 0, \quad C_{2,4} = \frac{1}{2}\left(\frac{m_*}{m^*}\right)^{1/2} - 1.
\end{aligned}
$$

Define $\{E_{k,q}\}_{(k,q)\in\mathbb{Z}_{\geq 3}\times\mathbb{Z}_{\geq 0}}$ by the recurrence relations

$$
\begin{aligned}
E_{3,q} &= \frac{1}{2}(m_*)^{-1/2}\Big\{-\sum_{j=3}^{q-1}(B_jE_{3,q+2-j} + E_{3,j}B_{q+2-j} + E_{3,j}E_{3,q+2-j})\\
&\quad + (C_{2,q-2} - B_{q-2}) + E_{4,q-2} - 2E_{3,q-2}\Big\} \text{ for } q \geq 5,\\
E_{a_k,q} &= (m_*)^{-1}\Big\{-\sum_{j=1}^{q-1}(A_jE_{a_k,q-j} + E_{a_k,j}A_{q-j} + E_{a_k,j}E_{a_k,q-j})\\
&\quad + (E_{a_k-1,q-4} + E_{a_k+1,q-4} - 2E_{a_k,q-4})\Big\} \text{ for } q \geq 5, k \geq 1,\\
E_{b_k,q} &= \frac{1}{2}(m_*)^{-1/2}\Big\{-\sum_{j=3}^{q-1}(B_jE_{b_k,q+2-j} + E_{b_k,j}B_{q+2-j} + E_{b_k,j}E_{b_k,q+2-j})\\
&\quad + (E_{b_k-1,q-2} + E_{b_k+1,q-2} - 2E_{b_k,q-2})\Big\} \text{ for } q \geq 5, k \geq 1,
\end{aligned}
$$

and the initial conditions

$$
\begin{aligned}
&E_{3,0} = E_{3,1} = E_{3,2} = E_{3,3} = 0, \quad E_{3,4} = -\frac{\sqrt{2}-1}{4\sqrt{2}},\\
&E_{k,0} = E_{k,1} = E_{k,2} = E_{k,3} = E_{k,4} = 0 \quad \text{for } k \geq 4.
\end{aligned}
$$

Finally, we define a mapping $k : \mathbb{Z}_{\geq 2} \to \mathbb{Z}_{\geq 2}$ by

$$k(\eta) = \begin{cases} 3d+1 & \text{if } \eta = 8d, 8d+1, \\ 3d+2 & \text{if } \eta = 8d+2, 8d+3, \\ 3d+3 & \text{if } \eta = 8d+4, 8d+5, 8d+6, 8d+7. \end{cases}$$

and its inverse by

$$d(k) := \min k^{-1}(\{k\}).$$

From the definition of $E_{k,q}$, we have

$$E_{k,q} = 0 \text{ for } 0 \leq q \leq d(k) - 1, \tag{7.21}$$

$$E_{a_k,d(a_k)} = (m_*)^{-1} E_{a_k-1,d(a_k-1)}, \tag{7.22}$$

$$E_{b_k,d(b_k)} = \frac{1}{2} (m_*)^{-1/2} E_{b_k-1,d(b_k-1)} \text{ for } k \geq 2. \tag{7.23}$$

We also prove that $u_{i,q}$ can be expressed by using A_q, B_q and $E_{k,q}$.

Proposition 6 *Let q be a non-negative integer. Choose $m \in \mathcal{M}^1_{3p+2}$ arbitrarily. If $0 \leq q \leq d(6r+3)$, the coefficients in the expansion of the positive solution of (1.3) are expressed as*

$$u_{i^r_-(a_k),q} = A_q + E_{a_k,q} + E_{6r+5-a_k,q} \quad \textit{for } a_k \in \mathbb{A}_r, \tag{7.24}$$

$$u_{i^r_-(b_k),q} = B_q + E_{b_k,q} + E_{6r+5-b_k,q} - 2\hat{E}_{q,r} \quad \textit{for } b_k \in \mathbb{B}_{2r}, \tag{7.25}$$

$$u_{i^r_+(a_k),q} = A_q + E_{a_k,q} + E_{6(p-1-r)+5-a_k,q} \quad \textit{for } a_k \in \mathbb{A}_{p-1-r}, \tag{7.26}$$

$$u_{i^r_+(b_k),q} = \begin{cases} B_q + E_{b_k,q} + E_{6(p-1-r)+5-b_k,q} & \textit{if } r < (p-1)/2, \\ B_q + E_{b_k,q} + E_{6(p-1-r)+5-b_k,q} - 2\hat{E}_{q,r} & \textit{if } r = (p-1)/2, \end{cases}$$

$$\textit{for } b_k \in \mathbb{B}_{2(p-1-r)}, \tag{7.27}$$

$$u_{i^r_-(2),q} = C_{2,q} + E_2, \tag{7.28}$$

$$u_{i^r_+(2),q} = \begin{cases} C_{2,q} & \textit{if } r < (p-1)/2, \\ C_{2,q} + E_2 & \textit{if } r = (p-1)/2, \end{cases} \tag{7.29}$$

$$u_{i^r_\pm(1),q} = C_{1,q}, \tag{7.30}$$

$$u_{i^r(0),q} = C_{0,q}, \tag{7.31}$$

where E_2 and $\hat{E}_{q,r}$ are constants given by

$$\hat{E}_2 = \begin{cases} 0 & \textit{if } 0 \leq q < d(6r+3), \\ (1/2)(m^*)^{-1} E_{6r+2,d(6r+2)} < 0 & \textit{if } q = d(6r+3), \end{cases}$$

$$\hat{E}_{q,r} = \begin{cases} 0 & \textit{if } 0 \leq q < d(6r+3), \\ E_{b_k,d(b_k)} E_{6r+5-b_k,d(6r+5-b_k)} > 0 & \textit{if } q = d(6r+3). \end{cases}$$

Proof The main idea of this proof is the same as Proposition 5. To prove the case $q = 0, 1, \ldots, 4$, we substitute every initial term of $u_{i,q}$ to (7.24)–(7.31) directly, which was determined in Sect. 6. Further, (7.24) may be proved in much the same way as the proof of (7.9). So we omit the proof.

To prove (7.25), we claim that for every $3 \leq j \leq d(6r+2)+1, 0 \leq q \leq d(6r+3)$ and $b_k \in \mathbb{B}_{2r}$, the following equalities hold:

$$
\begin{aligned}
&E_{b_k,j}E_{6r+5-b_k,q+2-j} = E_{6r+5-b_k,j}E_{b_k,q+2-j} \\
&\quad = \begin{cases} E_{b_k,d(b_k)}E_{6r+5-b_k,d(6r+5-b_k)} & \text{if } j = d(b_k), \\ 0 & \text{otherwise.} \end{cases}
\end{aligned} \tag{7.32}
$$

It is easy to verify that $d(6r+3)+2-d(b_k) = d(6r+5-b_k)$ for all $b_k \in \mathbb{B}_{2r}$. Hence if $q = d(6r+3)$ and $j = d(b_k)$, then the first two equalities in (7.32) hold. Otherwise, the proof of the equalities in (7.32) are t almost the same as that of (7.18), so we omit the proof. The rest steps of the proof of (7.25) is the same way as in that of (7.10), so we omit the proof. We can apply a similar argument to obtain (7.26) and (7.27). So we again omit the proof.

We prove (7.28) by induction on q. Suppose that (7.24)–(7.31) hold for $q = 1, 2, \ldots, s-1$, where s is a positive integer satisfying $s \leq d(6r+3)$. By (7.4), we have

$$
\begin{aligned}
u_{i^r_-(2),s} &= \frac{1}{2}\left(m^*\right)^{-1/2}\Bigg\{-\left(\sum_{j=3}^{s-1} C_{2,j}C_{2,s+2-j}\right) \\
&\quad + (B_{s-2} + E_{3,s-2} + E_{6r+2,s-2}) + C_{1,s-2} - 2C_{2,s-2}\Bigg\} \\
&= C_{2,s} + \frac{1}{2}\left(m^*\right)^{-1/2}(E_{6r+2,s-2}),
\end{aligned}
$$

In view of (7.21), $E_{6r+2,s-2} = 0$ for $4 \leq s-2 \leq d(6r+2)-1$. Further, we have $E_{6r+2,d(6r+2)} < 0$ by using (7.22)–(7.23) and the initial condition $E_{3,4}$. Hence we obtain

$$
u_{i^r_-(2),q} = C_{2,q} + \frac{1}{2}\left(m^*\right)^{-1/2}(E_{6r+2,d(6r+2)}) \quad \text{if } q = d(6r+3).
$$

Thus we have shown (7.28). The proof of (7.29)–(7.31) is almost the same as that of (7.28), so we omit the proof. $\square$

Let us complete the proof of Theorem 2. We consider $U(m^r_{3p+2}, \delta)$ for $1 \le r < p/2 - 1$. By Proposition 6 and (7.21), we have

$$
\begin{aligned}
U(m^r_{3p+2}, \delta) = & \sum_{q=0}^{d(6r+3)} \sum_{i=1}^{N} u_{i,q}\delta^{q/4} + o(\delta^{d(6r+3)/4}) \\
= & \sum_{q=0}^{3} (C_q + (p-1)(A_q + 2B_q)\delta^{q/4} \\
& + \sum_{q=4}^{d(6r+2)+1} (C_q + (p-1)(A_q + 2B_q) + \sum_{k=3}^{k(q)} 2E_{k,q})\delta^{q/4} \\
& + \sum_{q=d(6r+3)} \Big(C_q + E_2 + (p-1)(A_q + 2B_q) \\
& + \sum_{k=3}^{6r+2} 2E_{k,q} + E_{6r+3,q} - 4r\hat{E}_{q,r}\Big)\delta^{q/4} \\
& + o(\delta^{d(6r+3)/4}),
\end{aligned}
$$

where $C_q := C_{0,q} + 2C_{1,q} + 2C_{2,q}$. Hence we obtain

$$
\begin{aligned}
& U(m^{r+1}_{3p+2}, \delta) - U(m^r_{3p+2}, \delta) \\
& = (E_{6r+3,d(6r+3)} - E_2 + 4r\hat{E}_{q,r})\delta^{d(6r+3)/4} + o(\delta^{d(6r+3)/4}) \\
& = \left(-\frac{\sqrt{2}-1}{4\sqrt{2m_*}} E_{6r+2,d(6r+2)} + 4r\hat{E}_{q,r}\right)\delta^{d(6r+3)/4} + o(\delta^{d(6r+3)/4}) > 0
\end{aligned}
$$

for sufficiently small δ. This completes the proof of Theorem 2. □

8 Discussions

In this paper we studied a nonlinear optimization problem from population biology. We consider the population of a single species in a patchy environment and study the effects of dispersal and spatial heterogeneity of patches on the total population of a single species at equilibrium. More specifically, we ask the following question: Given the total amount of resources, how should the resources be distributed across the habitat in order to maximize the total population of a species? We show that the global maximizer can be characterized for any number of patches when the diffusion rate δ is either sufficiently small or large. Our results show that the global maximizer depends crucially on the diffusion rate δ, and the answers are completely different for small δ and large δ. In several cases we show that the global maximizer is of the "bang-bang" type, and we are also able to determine the maximizers explicitly by finding the specific guiding rules of fragmentation in the multi-patch model (1.3). In particular, fragmentation occurs when the diffusion rate is sufficiently small, which is in agreement with the findings in Mazari et al. (2020).

A general question is to determine the resource distributions which maximize the total population at equilibrium for the N-patch model

$$\frac{d}{dt}v_i(t) = v_i(m_i - v_i) + \delta \sum_{j=1}^{N} L_{ij}v_j, \tag{8.1}$$

where (L_{ij}) is non-negative, irreducible and

$$L_{ii} = -\sum_{j\neq i} L_{ij}, \qquad 1 \le i \le N.$$

For simplicity we assume that $L_{ij} = 1$ when patches i and j are connected, $L_{ij} = 0$ when they are disconnected. The patch model (1.1), which mimics the one-dimensional continuous habitat, has a special diffusion matrix, which shares similarity to the periodic case.

The answers to the above question might be complicated as both dispersal rate and dispersal matrix affect the mixing of populations across the whole habitat. The optimal resource distributions in patch models (in PDE models as well, but with an extra upper bound) are often of the bang-bang types, i.e. they are indicator functions over some set $E \subset \Omega$ or finite sums of indicator functions with different weights. The difficulty is to determine these sets E and their corresponding weights. Theorems 1, 2 and 3 provided some examples.

We suspect that Theorem 2 holds for small diffusion rate δ when $\eta = 0$. To be more precise, we conjecture that for model (1.1), there exists positive constant $\delta_{N,m} > 0$ such that $U(m, \delta) > U(\tilde{m}, \delta)$ holds for any $\delta \in (0, \delta_{N,m})$ and any $\tilde{m} \in \mathcal{M} \setminus \{m\}$. Note that the diffusion matrix given by model (1.1) is among the least connected dispersal matrices. Hence, this conjecture suggests that for small diffusion rate, in order to maximize the total population in weakly connected habitats, it might be advantageous to distribute the resources in certain fragmented manners, possibly so for model (8.1) as well.

On the other hand, Theorem 1 implies that it is advantageous to distribute the resources in a single patch when diffusion rate is large. It will be of interest to generalize Theorem 1 to model (8.1) for large δ and determine how the network topology affects the optimal distribution of resources. We suspect that the optimal distribution in this scenario might be associated with the boundary patches, i.e. patches only connected with a single patch.

If we increase the connectivity of the dispersal matrix, the optimal distribution of resources might also become less fragmented. To support this claim, consider the extreme case of completely connected habitat, i.e. $L_{ij} = 1$ for any $i \neq j$. For this case, it can be formally shown that for small diffusion rate, the optimal resource distribution is given by one of the following distributions:

$$\mathbf{m} = (m, 0, 0, 0, \ldots, 0),\ (0, m, 0, 0, \ldots, 0),\ \ldots\ (0, 0, 0, 0, \ldots, m),$$

with $U(\mathbf{m},\delta)$ given by

$$U(\mathbf{m},\delta)=m+(N-1)m^{1/2}\delta^{1/2}+o(\delta^{1/2}).$$

We wonder whether Theorem 3 can be extended to N-patch model with completely connected graphs.

Similarly for the PDE model (1.4), an open question is to characterize the global maximizer of the total population for the unique positive steady state in general domains. This seems to be a challenging problem even for the one-dimensional spatial domain with arbitrary dispersal rate.

A Appendix

A.1 Preliminaries for small δ

In this Appendix, we choose $N \geq 2$ and $m \in \mathcal{M}$ arbitrarily.

Lemma 7 *Any solution of* (1.3) *satisfies either* $\underline{\deg}\, u_i = 0$ *or* $1/2 \leq \underline{\deg}\, u_i < 1$ *for each* $i \in \Omega$. *Moreover,* $m_i > 0$ *if* $\underline{\deg}\, u_i = 0$, *and* $m_i = 0$ *if* $1/2 \leq \underline{\deg}\, u_i < 1$.

Proof By using (3.1)–(3.3), $u_i \to m_i$ as $\delta \to 0$. This implies $\overline{\deg}\, u_i \geq 0$. First, we choose $i = 1$ or $i = N$. Assume $m_i > 0$. Then by (3.1), we have

$$\begin{aligned} u_1 &= \frac{m_1-\delta}{2}+\left(\frac{1}{4}(m_1^2-2m_1\delta+\delta^2)+\delta u_2\right)^{1/2} \\ &= \frac{m_1-\delta}{2}+\frac{m_1}{2}\left(1+\frac{2}{m_1^2}\left(u_2-\frac{m_1}{2}\right)\delta+o(\delta)\right) \\ &= m_1-\left(\frac{u_2}{m_1}-1\right)\delta+o(\delta). \end{aligned}$$

Similarly, by (3.3), we have

$$u_N = m_N-\left(\frac{u_{N-1}}{m_N}-1\right)\delta+o(\delta).$$

This indicates that if $m_i > 0$, then $\overline{\deg}\, u_i = 0$.

Next, assume $m_i = 0$. By using (3.1) and (3.3), we have

$$u_1=-\frac{\delta}{2}+\left(\frac{1}{4}\delta^2+\delta u_2\right)^{1/2}, \quad u_N=-\frac{\delta}{2}+\left(\frac{1}{4}\delta^2+\delta u_{N-1}\right)^{1/2}.$$

This gives

$$\underline{\deg}\, u_1=\min\left\{1,\frac{1+\underline{\deg}\, u_2}{2}\right\}, \tag{A.1}$$

$$\underline{\deg}\, u_N = \min\left\{1, \frac{1+\underline{\deg}\, u_{N-1}}{2}\right\}. \tag{A.2}$$

This argument and Proposition 1 indicate that if $m_i = 0$, then $1/2 \le \underline{\deg}\, u_i \le 1$. It follows that $\underline{\deg}\, u_i = 0$ is equivalent to $m_i > 0$. From the above argument, we can assert that $1/2 \le \underline{\deg}\, u_i \le 1$ is equivalent to $m_i = 0$.

Second, we choose $i \in \mathbb{Z} \cap (1, N)$. Assume $m_i > 0$. By (3.2), we have

$$\begin{aligned} u_i &= \frac{m_i - 2\delta}{2} + \left(\frac{1}{4}m_i^2 - m_i\delta + \delta^2 + \delta(u_{i-1} + u_{i+1})\right)^{1/2} \\ &= \frac{m_i - 2\delta}{2} + \frac{m_i}{2}\left(1 + \frac{2}{m_i^2}(u_{i-1} + u_{i+1} - m_i)\delta + o(\delta)\right) \\ &= m_i + \left(\frac{u_{i-1} + u_{i+1}}{m_i} - 2\right)\delta + o(\delta). \end{aligned}$$

This implies that if $m_i > 0$, then $\underline{\deg}\, u_i = 0$. Next, assume $m_i = 0$. We use (3.2) again to have

$$u_i = -\delta + \left(\delta^2 + \delta(u_{i-1} + u_{i+1})\right)^{1/2}.$$

This gives

$$\underline{\deg}\, u_i = \min\left\{1, \frac{1 + \min\{\underline{\deg}\, u_{i-1}, \underline{\deg}\, u_{i+1}\}}{2}\right\}. \tag{A.3}$$

Similar argument to the first case shows that $\underline{\deg}\, u_i = 0$ is equivalent to $m_i > 0$. Furthermore, $1/2 \le \underline{\deg}\, u_i \le 1$ is equivalent to $m_i = 0$.

Therefore, it is sufficient to prove that $0 \le \underline{\deg}\, u_i < 1$ for all $i \in \Omega$. We assume that there exists $i \in \Omega$ such that $\underline{\deg}\, u_i = 1$. From (A.1)–(A.3), we have

$$\begin{cases} \underline{\deg}\, u_2 = 1 & \text{if } i = 1, \\ \underline{\deg}\, u_{i-1} = \underline{\deg}\, u_{i+1} = 1 & \text{if } 1 < i < N, \\ \underline{\deg}\, u_{N-1} = 1 & \text{if } i = N. \end{cases}$$

Repeated application of (A.1)–(A.3) enables us to obtain $m_i = 0$ for all $i \in \Omega$, which contradicts our assumption. □

Recall that (6.1)–(6.3) are obtained by Lemma 7. We next show that $\underline{\deg}\, u_i$ takes a finite number of values. By Lemma 8, we only have to seek the largest coefficients of an appropriate order of δ to maximize the total population.

Lemma 8 *Any solution of* (1.3) *satisfies*

$$\underline{\deg}\, u_i \in \left\{1 - 2^{-p} \mid p \in \mathbb{Z} \cap [0, N-1]\right\}$$

for all $i \in \Omega$.

Proof We assume that there exists $i \in \Omega$ such that

$$1 - 2^{-p} < \underline{\deg}\, u_i < 1 - 2^{-(p+1)}, \tag{A.4}$$

where $p \in \mathbb{Z} \cap [0, \min\{i-1, N-i\}]$. By using (A.1)–(A.3), we calculate

$$\begin{cases} \underline{\deg}\, u_1 = (1 + \underline{\deg}\, u_2)/2 & \text{if } i = 1, \\ \underline{\deg}\, u_i = (1 + \min\{\underline{\deg}\, u_{i-1}, \underline{\deg}\, u_{i+1}\})/2 & \text{if } i \in \mathbb{Z} \cap (1, N), \\ \underline{\deg}\, u_N = (1 + \underline{\deg}\, u_{N-1})/2 & \text{if } i = N. \end{cases} \tag{A.5}$$

Repeating substitution of (A.4) to (A.5) p times, we obtain

$$\begin{cases} 0 < \underline{\deg}\, u_{p+1} < 1/2 & \text{if } i = 1, \\ 0 < \min\{\underline{\deg}\, u_{i-p}, \underline{\deg}\, u_{i+p}\} < 1/2 & \text{if } i \in \mathbb{Z} \cap (1, N), \\ 0 < \underline{\deg}\, u_{N-p} < 1/2 & \text{if } i = N. \end{cases}$$

This contradicts Lemma 7. Finally, we assume that

$$1 - 2^{-(N-1)} < \underline{\deg}\, u_i.$$

Applying (A.1)–(A.3) repeatedly $\max\{N-i, i-1\}$ times to have $\min_{i\in\Omega}\{\underline{\deg}\, u_i\}$. Since $\max\{N-i, i-1\} \le N-1$, we have $\min_{i\in\Omega}\{\underline{\deg}\, u_i\} > 0$. This contradicts our assumption. □

We next evaluate the difference of $\underline{\deg}\, u_i$ between two adjacent patches.

Lemma 9 *Let $i \in \Omega$ and $p \in \mathbb{Z} \cap [0, N-1]$. Suppose that a solution of* (1.3) *satisfies*

$$\underline{\deg}\, u_i = \frac{2^p - 1}{2^p}.$$

Then we have

$$\frac{2^{p-1} - 1}{2^{p-1}} \le \min\{\underline{\deg}\, u_{i-1}, \underline{\deg}\, u_{i+1}\}, \tag{A.6}$$

where the equality holds if and only if $m_i = 0$, and

$$\max\{\underline{\deg}\, u_{i-1}, \underline{\deg}\, u_{i+1}\} \le \frac{2^{p+1} - 1}{2^{p+1}}, \tag{A.7}$$

where the equality holds if and only if $\min\{m_{i-1}, m_{i+1}\} = 0$.

Proof It is clear that (A.6) follows immediately from (A.5). Suppose that

$$\max\{\underline{\deg}\, u_{i-1}, \underline{\deg}\, u_{i+1}\} \ge \frac{2^{p+2} - 1}{2^{p+2}}.$$

By (A.6), we have

$$\min\{\underline{\deg}\, u_{i-2}, \underline{\deg}\, u_i\} = \frac{2^{p+1}-1}{2^{p+1}}, \text{ or } \min\{\underline{\deg}\, u_i, \underline{\deg}\, u_{i+2}\} = \frac{2^{p+1}-1}{2^{p+1}}.$$

This contradicts our assumption. This proves (A.7). □

Repeated application of Lemma 9 enables us to have the following lemma.

Lemma 10 *Let $i, k \in \Omega$, and take $p \in \mathbb{Z} \cap [1, N-1]$ such that $k-p, k+p \in \Omega$. If a positive solution of (1.3) satisfies $\underline{\deg}\, u_k = (2^p-1)/2^p$, then*

$$\min\{\underline{\deg}\, u_{\max\{1,k-p\}}, \underline{\deg}\, u_{\min\{k+p,N\}}\} = 0.$$

Moreover, $m_i = 0$ for all $i \in (k-p, k+p)$.

Lemma 10 allows us to examine the effect of the distance between favorable patches on $\underline{\deg}\, u_i$. The following proposition plays an important role in Sect. 6.

Lemma 11 *Let k, i, q be positive integers such that $k, k+i, k+2q \in \Omega$.*

(i) Suppose that $m \in \mathcal{M}$ satisfies $m_k > 0$, $m_{k+2q} > 0$, and $m_{k+i} = 0$ $(0 < i < 2q)$. Then the following equalities hold:

$$\underline{\deg}\, u_{k+i} = \underline{\deg}\, u_{k+2q-i} = \frac{2^i-1}{2^i} \quad (0 \le i \le q).$$

(ii) Suppose that $m \in \mathcal{M}$ satisfies $m_k > 0$, $m_{k+2q} > 0$, and $m_{k+i} = 0$ $(0 < i < 2q+1)$. Then the following equalities hold:

$$\underline{\deg}\, u_{k+i} = \underline{\deg}\, u_{k+2q+1-i} = \frac{2^i-1}{2^i} \quad (0 \le i \le q).$$

Proof (i) Assume that $i \in \mathbb{Z} \cap [1, q]$ and

$$\underline{\deg}\, u_{k+i} = \frac{2^{i-j}-1}{2^{i-j}} < (2^i-1)/2^i,$$

where $j \in \mathbb{Z} \cap [1, i]$. Then we have

$$\min\{\underline{\deg}\, u_{k+j}, \underline{\deg}\, u_{k+2i-j}\} = 0$$

by Lemma 10. Similarly, assuming that

$$\underline{\deg}\, u_{k+i} = \frac{2^{i+j}-1}{2^{i+j}} > (2^i-1)/2^i,$$

where $j \in \mathbb{Z} \cap [1, N-1-i]$, we have $m_k = 0$ by Lemma 10. This contradicts our assumption $m_k > 0$. We now apply this argument again, with $k+i$ replaced by $k+2q-i$, to obtain the value of $\underline{\deg}\, u_{k+2q-i}$.

The case (ii) may be proved in much the same way as the case (i), so we omit the proof. □

A.2 Expansion of the total population in *N* patches

To express the total population $U(m, \delta)$ in terms of δ for each $m \in \mathcal{M}$, we show the following result.

Proposition 7 *Let k, i, q be positive integers such that $k, k+i, k+2q \in \Omega$. Choose $m \in \mathcal{M}$ arbitrarily. Then the positive solution of (1.3) are expanded as follows:*

(i) *If $m \in \mathcal{M}$ satisfies $m_k > 0$, $m_{k+2q} > 0$ and $m_{k+i} = 0$, $(1 \le i \le 2q-1)$, then*

$$u_k = \begin{cases} m_k - \delta + o(\delta), & (k=1), \\ m_k + \left(\dfrac{m_{k-1}}{m_k} - 2\right)\delta + o(\delta), & (k>1), \end{cases} \tag{A.8}$$

$$u_{k+i} = m_k^{1/2^i}\delta^{(2^i-1)/2^{-i}} - \delta + o(\delta), \ (1 \le i \le q-1), \tag{A.9}$$

$$u_{k+q} = \sqrt{m_k^{1/2^{q-1}} + m_{k+2q}^{1/2^{q-1}}}\,\delta^{(2^q-1)/2^q} - \delta + o(\delta), \tag{A.10}$$

$$u_{k+2q-i} = m_{k+2q}^{1/2^i}\delta^{(2^i-1)/2^i} - \delta + o(\delta), \ (1 \le i \le q-1), \tag{A.11}$$

$$u_{k+2q} = \begin{cases} m_{k+2q} - \delta + o(\delta), & (k+2q=N), \\ m_{k+2q} + \left(\dfrac{m_{k+2q+1}}{m_{k+2q}} - 2\right)\delta + o(\delta), & (k+2q<N). \end{cases} \tag{A.12}$$

(ii) *If $m \in \mathcal{M}$ satisfies $m_k > 0$, $m_{k+2q+1} > 0$ and $m_{k+i} = 0$, $(1 \le i \le 2q)$, then*

$$\begin{aligned}
u_k &= \begin{cases} m_k - \delta + o(\delta), & (k=1), \\ m_k + \left(\dfrac{m_{k-1}}{m_k} - 2\right)\delta + o(\delta), & (k>1), \end{cases} \\
u_{k+i} &= m_k^{1/2^i}\delta^{(2^i-1)/2^i} - \delta + o(\delta), \ (1 \le i \le q-1), \\
u_{k+q} &= m_k^{1/2^q}\delta^{(2^q-1)/2^q} + \left(\frac{1}{2}\left(\frac{m_{k+2q+1}}{m_k}\right)^{1/2^q} - 1\right)\delta + o(\delta), \qquad \text{(A.13)} \\
u_{k+q+1} &= m_{k+2q+1}^{1/2^q}\delta^{(2^q-1)/2^q} + \left(\frac{1}{2}\left(\frac{m_k}{m_{k+2q+1}}\right)^{1/2^q} - 1\right)\delta + o(\delta), \\
u_{k+2q+1-i} &= m_{k+2q+1}^{1/2^i}\delta^{(2^i-1)/2^i} - \delta + o(\delta), \ (1 \le i \le q-1), \\
u_{k+2q+1} &= \begin{cases} m_{k+2q+1} - \delta + o(\delta), & (k+2q+1=N), \\ m_{k+2q+1} + \left(\dfrac{m_{k+2q+2}}{m_{k+2q+1}} - 2\right)\delta + o(\delta), & (k+2q+1<N). \end{cases} \qquad \text{(A.14)}
\end{aligned}$$

(iii) *If* $m \in \mathcal{M}$ *satisfies* $m_k > 0,\ (k > 1),\ m_i = 0,\ (1 \le i \le k-1)$, *then*

$$\begin{aligned} u_1 &= m_k^{1/2^{k-1}} \delta^{(2^{k-1}-1)/2^{k-1}} - \frac{\delta}{2} + o(\delta), \\ u_{k-i} &= m_k^{1/2^i} \delta^{(2^i-1)/2^i} - \delta + o(\delta), \\ u_k &= \begin{cases} m_k - \delta + o(\delta), & (k = N), \\ m_k + \left(\dfrac{m_{k+1}}{m_k} - 2 \right) \delta + o(\delta), & (k < N), \end{cases} \end{aligned} \tag{A.15}$$

(iv) *If* $m \in \mathcal{M}$ *satisfies* $m_k > 0,\ (k < N),\ m_i = 0,\ (k+1 \le i \le N)$, *then*

$$\begin{aligned} u_k &= \begin{cases} m_k - \delta + o(\delta), & (k = 1), \\ m_k + \left(\dfrac{m_{k-1}}{m_k} - 2 \right) \delta + o(\delta), & (k > 1), \end{cases} \\ u_{k+i} &= m_k^{1/2^i} \delta^{(2^i-1)/2^i} - \delta + o(\delta), \\ u_N &= m_k^{1/2^{N-k}} \delta^{(2^{N-k}-1)/2^{N-k}} - \frac{\delta}{2} + o(\delta). \end{aligned}$$

Proof (i) It is easy to verify (A.8) by noting $\underline{\deg}\, u_{k+1} = 1/2$. So we omit the proof. We now consider (A.9) by induction on i. By Lemma 11, there exist coefficients C_{k+i} for all $i \in \{0, 1, \dots, q\}$ such that

$$u_{k+i} = C_{k+i} \delta^{(2^i-1)/2^i} + o(\delta^{(2^i-1)/2^i}), \quad C_k = m_k.$$

By (6.1) or (6.2), we have the base case as

$$\begin{aligned} u_{k+1} &= \left(m_k \delta + C_{k+2} \delta^{7/4} + o(\delta^{7/4}) \right)^{1/2} - \delta + o(\delta) \\ &= (m_k \delta)^{1/2} \left(1 + \frac{C_{k+2}}{2m_k} \delta^{3/4} + o(\delta^{3/4}) \right) - \delta + o(\delta) \\ &= m_k^{1/2} \delta^{1/2} - \delta + o(\delta). \end{aligned}$$

Therefore, we have the relation between C_k and C_{k+1} given by

$$C_{k+1} = C_k^{1/2}.$$

Suppose that

$$u_{k+i-1} = C_{k+i-1} \delta^{(2^{i-1}-1)/2^{i-1}} - \delta + o(\delta) \quad (3 \le i \le q-1).$$

By (6.2), we have the induction step as

$$
\begin{aligned}
u_{k+i} &= \left(C_{k+i-1}\delta^{(2^i-1)/2^{i-1}} + C_{k+i+1}\delta^{(2^{i+2}-1)/2^{i+1}} + o(\delta^{(2^{i+2}-1)/2^{i+1}})\right)^{1/2} - \delta + o(\delta) \\
&= C_{k+i-1}^{1/2}\delta^{(2^i-1)/2^i}\left(1 + \frac{C_{k+i+1}}{2C_{k+i-1}}\delta^{1/2^{i-1}-1/2^{i+1}} + o(\delta^{1/2^{i-1}-1/2^{i+1}})\right) - \delta + o(\delta) \\
&= C_{k+i-1}^{1/2}\delta^{(2^i-1)/2^i} - \delta + o(\delta).
\end{aligned}
$$

Therefore, we have recurrence relation $C_{k+i} = C_{k+i-1}^{1/2}$ which gives (A.9). From this, we replace "k" by "$k+2q$" to have (A.11) and (A.12).

To prove (A.10), we use (A.9) and (A.11) to have

$$
\begin{aligned}
u_{k+q-1} &= m_k^{1/2^{q-1}}\delta^{(2^{q-1}-1)/2^{q-1}} - \delta + o(\delta), \\
u_{k+q+1} &= m_{k+2q}^{1/2^{q-1}}\delta^{(2^{q-1}-1)/2^{q-1}} - \delta + o(\delta).
\end{aligned}
$$

By (6.2), we obtain

$$
\begin{aligned}
u_{k+q} &= \left(m_k^{1/2^{q-1}}\delta^{(2^q-1)/2^{q-1}} + m_{k+2q}^{1/2^{q-1}}\delta^{(2^q-1)/2^{q-1}} - 2\delta^2 + o(\delta^2)\right)^{1/2} - \delta + o(\delta) \\
&= \left(m_k^{1/2^{q-1}} + m_{k+2q}^{1/2^{q-1}}\right)^{1/2}\delta^{(2^q-1)/2^q}\left(1 - \frac{\delta^{1/2^{q-1}}}{m_k^{1/2^{q-1}} + m_{k+2q}^{1/2^{q-1}}} + o(\delta^{1/2^{q-1}})\right) \\
&\quad - \delta + o(\delta) \\
&= \left(m_k^{1/2^{q-1}} + m_{k+2q}^{1/2^{q-1}}\right)^{1/2}\delta^{(2^q-1)/2^q} - \delta + o(\delta).
\end{aligned}
$$

(ii) We give the proof only for (A.13) and (A.14); the other cases can be proved by the same argument as above. We use Lemma 11 to have

$$
\begin{aligned}
u_{k+q} &= C_{k+q}\delta^{(2^q-1)/2^q} + o(\delta^{(2^q-1)/2^q}), \\
u_{k+q+1} &= C_{k+q+1}\delta^{2^q-1/2^q} + o(\delta^{(2^q-1)/2^q}).
\end{aligned}
$$

We now compute C_{k+q} and C_{k+q+1} by using (6.2) as

$$
\begin{aligned}
u_{k+q} &= \left(m_k^{1/2^{q-1}}\delta^{(2^q-1)/2^{q-1}} + C_{k+q+1}\delta^{(2^{q+1}-1)/2^q} + o(\delta^{(2^{q+1}-1)/2^q})\right)^{1/2} - \delta + (\delta) \\
&= m_k^{1/2^q}\delta^{(2^q-1)/2^q}\left(1 + \frac{C_{k+q+1}}{2m_k^{1/2^{q-1}}}\delta^{1/2^q} + o(\delta^{1/2^q})\right) - \delta + (\delta) \\
&= m_k^{1/2^q}\delta^{(2^q-1)/2^q} + \left(\frac{C_{k+q+1}}{2m_k^{1/2^q}} - 1\right)\delta + (\delta).
\end{aligned}
$$

Similarly, we have

$$u_{k+q+1} = m_{k+2q+1}^{1/2^q} \delta^{(2^q-1)/2^q} + \left(\frac{C_{k+q}}{2m_{k+2q+1}^{1/2^q}} - 1 \right) \delta + (\delta).$$

By the definition of C_{k+q} and C_{k+q+1}, we have

$$C_{k+q} = m_k^{1/2^q}, \quad C_{k+q+1} = m_{k+2q+1}^{1/2^q}.$$

Therefore, we obtain (A.13) and (A.14).
(iii) We use (6.1) to have

$$\begin{aligned}
u_1 &= \left(\delta \left(m_k^{1/2^{k-2}} \delta^{(2^{k-2}-1)/2^{k-2}} - \delta + o(\delta) \right) \right)^{1/2} - \frac{\delta}{2} + o(\delta) \\
&= m_k^{1/2^{k-1}} \delta^{(2^{k-1}-1)/2^{k-1}} \left(1 - \frac{1}{2m_k^{1/2^{k-2}}} \delta^{1/2^{k-2}} + o(\delta^{1/2^{k-2}}) \right) - \frac{\delta}{2} + o(\delta) \\
&= m_k^{1/2^{k-1}} \delta^{(2^{k-1}-1)/2^{k-1}} - \frac{\delta}{2} + o(\delta).
\end{aligned}$$

This proves (A.15) The other cases can be proved by the same argument.
(iv) This can be proved in much the same way as (iii). So we omit the proof. □

References

Allen LJS (1987) Persistence, extinction, and critical patch number for island populations. J Math Biol 24:617–625

Bai XL, He XQ, Li F (2016) An optimization problem and its application in population dynamics. Proc Am Math Soc 144:2161–2170

Cantrell RS, Cosner C (2003) Spatial ecology via reaction-diffusion equations. Series in mathematical and computational biology. Wiley, Chichester

Cantrell RS, Cosner C, Hutson V (1996) Ecological models, permanence and spatial heterogeneity. Rocky Mountain J Math 26:1–35

Ding W, Finotti H, Lenhart S, Lou Y, Ye Q (2010) Optimal control of growth coefficient on a steady-state population model. Nonlinear Anal Real World Appl 11:688–704

Hastings A (1982) Dynamics of a single species in a spatially varying environment: the stabilizing role of high dispersal rates. J Math Biol 16:49–55

Hirsch MW (1984) The dynamical systems approach to differential equations. Bull Am Math Soc 11(1):1–64

Holt RD (1985) Population dynamics in two-patch environments: some anomalous consequences of an optimal habitat distribution. Theor Popul Biol 28:181–208

Levin SA (1974) Dispersion and population interactions. Am Nat 108(960):207–228

Levins R (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. Am Entomol 15(3):237–240

Lou Y (2006) On the effects of migration and spatial heterogeneity on single and multiple species. J Differ Equ 223:400–426

Lou Y, Yanagida E (2006) Minimization of the principal eigenvalue for an elliptic boundary value problem with indefinite weight and applications to population dynamics. Japan J Ind Appl Math 23:275–292

Mazari I, Ruiz-Balet D (2020) A fragmentation phenomenon for a non-energetic optimal control problem: optimisation of the total population size in logistic diffusive models. SIAM J Appl Math

Mazari I, Nadin G, Privat Y (2020) Optimal location of resources maximizing the total population size in logistic models. J de Math Pures et Appliquees 134:1–35

Mazari I, Nadin G, Privat Y (2020) Shape optimization of a weighted two-phase Dirichlet eigenvalue, ArXiv preprint

Nagahara K, Yanagida E (2018) Maximization of the total population in a reaction-diffusion model with logistic growth. Calculus Variations Part Differ Equ 57(3):80

Skellam JG (1951) Random dispersal in theoretical populations. Biometrika 38:196–218

Takeuchi Y (1989) Cooperative systems theory and global stability of diffusion models. Acta Appl Math 14:49–57