The Annals of Applied Probability
2005, Vol. 15, No. 3, 2038–2061
DOI 10.1214/105051605000000368

STRONG APPROXIMATION FOR THE SUPERMARKET MODEL

BY MALWINA J. LUCZAK AND JAMES NORRIS

London School of Economics and University of Cambridge

We prove three strong approximation theorems for the "supermarket" or "join the shortest queue" model—a law of large numbers, a jump process approximation and a central limit theorem. The estimates are carried through rather explicitly, and rely in part on couplings. This allows us to approximate each of the infinitely many components of the process in its own scale and to exhibit a cut-off in the set of active components which grows slowly with the number of servers.

1. Introduction. The supermarket model is a system of N single-server queues. Customers arrive as a Poisson process of rate $N\lambda$, where $\lambda \in (0, 1)$. Each customer examines d queues, chosen randomly from all N queues, where $d \geq 2$, and joins the shortest of these d queues, choosing randomly if the shortest queue is not unique. The service times of all customers are independent rate 1 exponential random variables. We will be concerned with the behavior of this model when λ and d are fixed, over a finite time interval $[0, t_0]$, as $N \to \infty$. We shall consider the case when the system starts in some well-behaved state with low server loads (in a sense to be made precise below).

This model has attracted attention because it turns out that the choice offered to customers, even if $d = 2$, dramatically reduces queue lengths (see [5, 14, 16]) and, in particular, the length of the longest queue (see [10, 11]). Given that our analysis relies on N being very much larger than d, the model does not describe well the behavior of a real supermarket. Rather it serves as an example where a simple dynamic routing rule leads to a greatly improved performance, which is of interest in the context of communications networks.

Our results provide strong approximations for the supermarket model, and include a law of large numbers and a diffusion approximation. In arriving at these results, we have developed techniques to establish weak convergence of a sequence of Markov processes X^N in infinitely many dimensions, where the jumps of X^N are of order N^{-1} and occur at a rate of order N. The classical results for fluid limits are set in a finite-dimensional context. We make essential use of the fact that the number of "active" components in X^N grows only very slowly with N. We have used direct and quantitative methods based on exponential

Received May 2004; revised December 2004.
AMS 2000 subject classifications. Primary 60K25; secondary 60F15.
Key words and phrases. Supermarket model, join the shortest queue, law of large numbers, diffusion approximation, exponential martingale inequalities, coupling.

martingales and strong approximation of Poisson processes by Brownian motion. These methods seem well-suited to deal with such “almost finite-dimensional” Markov processes. Earlier results for this model include laws of large numbers in [5, 6, 14, 16], quantitative concentration of measure estimates in [10] and a central limit theorem [7]. See also [9] for a preliminary version of the law of large numbers presented in the present paper (for a more general range of initial conditions).

The limiting behavior of the supermarket model as $N \to \infty$ may conveniently be described in terms of the vector $X_t = (X_t^k : k \in \mathbb{N})$, where X_t^k denotes the proportion of all N queues having at least k customers at time t. The process $X = (X_t)_{t \geq 0}$ is a Markov chain. We will suppose throughout that $X_0 = x_0$ with x_0 nonrandom and we will suppress the dependence of X on N to lighten the notation. Now X has the form of a density dependent Markov chain such as considered by Ethier and Kurtz in [4], Chapter 11. Thus, one might expect to be able to find a deterministic process $(x_t)_{t \geq 0}$ and a Gaussian process $(\gamma_t)_{t \geq 0}$ such that

$$X_t = x_t + O(N^{-1/2}), \qquad X_t = x_t + N^{-1/2}\gamma_t + O(\log N/N).$$

However, the number of nonzero components in X grows with N so the standard theory does not apply.

We will see that, for small initial data, the component X^k has a scale $a_k = \lambda^{1+d+\cdots+d^{k-1}}$, which, of course, decays very rapidly with k. Thus, the number of queues having at least k customers is of order Na_k. We can find m of order $\log\log N$ such that Na_m is of order 1. Thus, we can exhibit a cut-off in the number of active components which grows only slowly with N. Below the cut-off, for $k \leq m-1$, we prove convergence with explicit control of error probabilities for each of the $\log\log N$ active components.

We thereby obtain results of the form

$$X_t^k = x_t^k + a_k O\big(\sqrt{\log\log\log N/Na_k}\big),$$
$$X_t^k = x_t^k + N^{-1/2}\gamma_t^k + a_k O\big(\log(Na_k)/Na_k\big).$$

Note that each component is estimated in the correct scale, with an error depending on the number of queues active at that level. The $\log\log\log N$ in the first equation is a (small) price we pay for working with infinitely many components. These asymptotics will be established with a degree of uniformity in x_0, which thus allows a dependence of x_0 on N. The Gaussian approximation relies, as in the finite-dimensional case, on a sophisticated coupling of the compensated Poisson process with Brownian motion due to Komlós, Major and Tusnády [8].

We will give a third result, for $k \leq m-1$, of the form

$$X_t^k = x_t^k + N^{-1/2}\tilde{\gamma}_t^k + a_k O\big((\log\log\log N/Na_k)^{3/4}\big).$$

Here $(\tilde{\gamma}_t)_{t \geq 0}$ is a jump process with drift which depends on N but is of a simpler type than X in that it is a linear function of additive Poisson noise. The characteristics of $\tilde{\gamma}$ are derived in a simple and canonical way from those of X. Moreover, $\tilde{\gamma}$ and X share a common filtration. The error term is larger than for the Gaussian approximation. On the other hand, the derivation is significantly simpler.

We obtain also the behavior of the queue sizes at and above the cut-off. We see a residual randomness in $(X_t^m)_{t\geq 0}$ even for large values of N. This may be approximated in terms of an $M/M/\infty$ queue with arrival rate $N\lambda(x_t^{m-1})^d$ and service rate 1. Over a given finite time interval, there are no queues with lengths greater than m.

Thus, we will show for the supermarket model that its infinite-dimensional character does not prevent the derivation of precise asymptotics. We expect the general approach taken here to adapt well to a number of further examples of similar character.

2. Statement of results. Let S_0 denote the set of nonincreasing sequences $x = (x^k : k \in \mathbb{N})$ in $[0, 1]$, where $\mathbb{N} = \{1, 2, \dots\}$. For $x \in S_0$, set $x^0 = 1$ and define $\lambda_+^k(x) = \lambda((x^{k-1})^d - (x^k)^d)$, $\lambda_-^k(x) = x^k - x^{k+1}$ and $b^k(x) = \lambda_+^k(x) - \lambda_-^k(x)$. It is shown in [16] that, given $x_0 \in S_0$, there is a unique solution $(x_t)_{t\geq 0}$ to $\dot{x}_t = b(x_t)$ in S_0. Moreover, for any other solution $(y_t)_{t\geq 0}$ in S_0, $x_0^k \leq y_0^k$ for all k implies $x_t^k \leq y_t^k$ for all k and all $t \geq 0$.

Recall that $a_k = \lambda^{1+d+\cdots+d^{k-1}}$ for $k \in \mathbb{N}$. Then $a = (a_k : k \in \mathbb{N})$ is the unique solution in S_0 to $b(x) = 0$ such that $\lim_{k\to\infty} x^k = 0$; to see this note that $x^{k+1} - \lambda(x^k)^d$ is independent of k for any solution x.

Define $\|x\| = \sup_k |x^k|/a_k$, set $E = \{x \in \mathbb{R}^{\mathbb{N}} : \|x\| < \infty\}$ and set $S = S_0 \cap E$. We shall work on a fixed interval $[0, t_0]$. For us, the good initial conditions will be those $x_0 \in S$ for which $\|x_t\| \leq \rho$ for all $t \in [0, t_0]$, for some $\rho < \infty$. Write $S(\rho, t_0)$ for the set of such x_0. We note that, if $x_0 \in S_0$ with $\|x_0\| \leq 1$, then, by comparison with the stationary solution a, $\|x_t\| \leq 1$ for all t, so $x_0 \in S(1, t_0)$ for all t_0. More generally, if $x_0 \in S$, then $t \mapsto \|x_t\|$ is continuous (and finite) on some interval $[0, \zeta)$, so $x_0 \in S(\rho, t_0)$ for some $\rho < \infty$ for all $t_0 < \zeta$. To see this, we extend $\lambda_\pm$ and b to $\mathbb{R}^{\mathbb{N}}$ by setting $\lambda_+^k(x) = \lambda((y^{k-1})^d - (y^k)^d)^+$, $\lambda_-^k(x) = (y^k - y^{k+1})^+$, where $y^k = (x^k)^+$; then b maps E to itself and is locally Lipschitz for the given norm [to prove this fact, carry out estimates similar to those in (5) and (6) below]; so if $x_0 \in S$, then there is a local solution $(x_t)_{t<\zeta}$ in S, which must coincide with the global solution $(x_t)_{t\geq 0}$ in S_0 for $t < \zeta$.

The state-space I of the Markov chain $X = (X_t)_{t\geq 0}$ is the set of nonincreasing sequences in $N^{-1}\{0, 1, \dots, N\}$ with finitely many nonzero terms. Thus, $I \subseteq S$. The Lévy kernel for X is given by

$$K(x, dy) = \sum_{k=1}^{\infty} \big[N\lambda_+^k(x)\delta_{e_k/N}(dy) + N\lambda_-^k(x)\delta_{-e_k/N}(dy)\big],$$

where e_k denotes the kth standard basis vector. Given $m \in \mathbb{N}$, let $(\hat{X}_t^m)_{t \le t_0}$ be a process starting from x_0^m and such that $N\hat{X}^m$ is an $M/M/\infty$ queue, with arrival rate $N\lambda(x_t^{m-1})^d$ and service rate 1. We can now state our law of large numbers.

THEOREM 2.1. *Set $m = m(N) = \inf\{k \in \mathbb{N} : Na_k \le (\log N)^4\}$. There is a coupling of $\hat{X}^m$ and X^m such that, for all $\rho \ge 1, t_0 > 0$ and all sequences $R(N)$ with $R(N)/\sqrt{\log\log\log N} \to \infty$, we have*

$$\sup_{x_0 \in I \cap S(\rho, t_0)} \mathbb{P}_{x_0}\big(\sqrt{N}|X_t^{N,k} - x_t^k| > R(N)\sqrt{a_k} \text{ for some } k \le m-1$$
$$\text{or } X_t^{N,m} \ne \hat{X}_t^{N,m} \text{ or } X_t^{N,m+1} \ne 0 \text{ for some } t \le t_0\big) \to 0.$$

In particular,

$$\sup_{k \le m-1} \sup_{t \le t_0} |X_t^{k,N} - x_t^k|/a_k \to 0$$

in probability, uniformly in $x_0 \in I \cap S(\rho, t_0)$.

We know (see [16]) that if $\rho > 0$ and $\|x_0\| \le \rho$, then $x_t^k \to a_k$ as $t \to \infty$. Thus, for $k \le m-1$, the proportional error in approximating X_t^k by the deterministic process x_t^k is small for large values of N.

The central limit theorem shows generically that the power $\sqrt{N}$ in Theorem 2.1 cannot be improved while the approximating process $(x_t)_{t \ge 0}$ remains deterministic. Our next result is a refined approximation which allows an improvement to $N^{3/4}$. Let $\tilde{\mu}$ be a Poisson random measure on $\mathbb{R}^{\mathbb{N}} \times (0, t_0]$ with intensity

$$\tilde{\nu}(dy, dt) = K(x_t, dy)\, dt.$$

For any y,

$$\nabla b^k(x) y = \lambda d (x^{k-1})^{d-1} y^{k-1} - \lambda d (x^k)^{d-1} y^k - y^k + y^{k+1}.$$

We show in Section 6 that the linear equations

$$\tilde{\gamma}_t^k = \sqrt{N} \int_{\mathbb{R}^{\mathbb{N}} \times (0,t]} y^k (\tilde{\mu} - \tilde{\nu})(dy, ds) + \int_0^t \nabla b^k(x_s) \tilde{\gamma}_s\, ds \tag{1}$$

have a unique cadlag solution $(\tilde{\gamma}_t^k : k \in \mathbb{N}, t \le t_0)$. Set $\tilde{X}_t = x_t + N^{-1/2} \tilde{\gamma}_t$.

THEOREM 2.2. *Define $m(N)$ as in Theorem 2.1. There is a coupling of $\tilde{X}$ and X, in a common filtration, such that, for all $\rho \ge 1$, $t_0 > 0$ and all sequences $\tilde{R}(N)$ with $\tilde{R}(N)/(\log\log\log N)^{3/4} \to \infty$, we have*

$$\sup_{x_0 \in I \cap S(\rho, t_0)} \mathbb{P}_{x_0}\big(N^{3/4}|X_t^k - \tilde{X}_t^k| > \tilde{R}(N) a_k^{1/4} \text{ for some } k \le m-1, t \le t_0\big) \to 0.$$

The final result is a diffusion approximation. Let B_+^k, B_-^k, $k \in \mathbb{N}$, be independent standard Brownian motions. Set $\sigma_\pm^k(x) = \sqrt{\lambda_\pm^k(x)}$. We show in Section 6 that the linear equations

$$\gamma_t^k = \int_0^t \sigma_+^k(x_s)\,dB_+^k(s) - \int_0^t \sigma_-^k(x_s)\,dB_-^k(s) + \int_0^t \nabla b^k(x_s)\gamma_s\,ds, \qquad t \le t_0, \tag{2}$$

have a unique solution $(\gamma_t^k : k \in \mathbb{N}, t \le t_0)$ with

$$\sup_{k\in\mathbb{N}} \mathbb{E}\Big(\sup_{t\le t_0} |\gamma_t^k|^2\Big) < \infty.$$

Set $\bar{X}_t = x_t + N^{-1/2}\gamma_t$.

THEOREM 2.3. *Define $m(N)$ as in Theorem 2.1. There is a coupling of $\bar{X}$ and X such that, for all $\rho \ge 1$, $t_0 > 0$, there is a constant $\bar{R}$, independent of N, such that*

$$\sup_{x_0\in I\cap S(\rho,t_0)} \mathbb{P}_{x_0}\big(N|X_t^k - \bar{X}_t^k| > \bar{R}\log(Na_k) \text{ for some } k \le m-1, t \le t_0\big) \to 0.$$

We remark that there are alternative versions of all three theorems in which x_0 is replaced by $x_0^{(m)} = (x_0^1, \dots, x_0^{m-1}, 0, \dots)$ and $\lambda_\pm(x)$ is replaced by $\lambda_\pm(x)^{(m)}$, so that the approximating deterministic dynamics are $(m-1)$-dimensional. The proofs are a minor modification of the proofs given below. These alternative versions would have merit in any computational implementation of the approximations since m is of order only $\log\log N$.

In comparison with previous results, our theorems for the first time approximate each component of the infinite dimensional process in its own scale, while at the same time providing explicit rates of convergence. In particular, Theorem 2.1 strengthens the law of large numbers in [5, 6, 16], and Theorem 2.3 strengthens the central limit theorem in [7]. Unlike the techniques developed in [10], ours apply only on finite time intervals and do not extend to the equilibrium distribution. On the other hand, the estimates in [10] do not distinguish between the magnitudes of different components of the process.

3. Law of large numbers. In the first half of this section, we fix N and $A, R \ge 1$ and set $m = \inf\{k \in \mathbb{N} : Na_k \le A\}$. We will obtain, subject to certain constraints, a global estimate on the probability appearing in Theorem 2.1. In the second half we will show that this estimate becomes small as $N \to \infty$ when $A = (\log N)^4$ and when R is chosen as in Theorem 2.1.

The integer m, which will turn out to give the maximum queue length, is of order $\log\log N/\log d$ for the values of A we shall consider. To see this, fix $\rho \ge 1$ and

$t_0 > 0$ and assume that $\rho A^d \le N^{d-1}$. Set $\alpha = (\log\log N - \log\log(1/\lambda))/\log d$, so that $N\lambda^{d^\alpha} = 1$ and $Na_k = \lambda^{1+d+\cdots+d^{k-1}-d^\alpha}$. If $k \le \alpha$, then $Na_k \ge 1$, whereas if $k \ge \alpha + 1$, then $Na_k \le 1$. Hence, at least for sufficiently large N, we will have $m \in (\alpha - 2, \alpha + 2)$.

Consider the case $x_0 \in S(\rho, t_0)$. Then $Nx_0^{m+1} \le N\rho a_{m+1} < N\rho a_m^d \le \rho A^d / N^{d-1} \le 1$, so $x_0^{m+1} = 0$. Set $T_1 = \inf\{t \ge 0 : X_t^{m+1} \ne 0\}$. Note that, while $X_t^{m+1} = 0$, X_t^{m+1} increases at rate $N\lambda(X_{t-}^m)^d$, whereas $X_t^m + X_t^{m+1}$ increases at rate $N\lambda(X_{t-}^{m-1})^d$ and decreases at rate NX_{t-}^m. We can therefore find an $M/M/\infty$ queue $(Q_t)_{t\ge 0}$, starting from Nx_0^m, without arrivals and with service rate 1, and a Poisson random measure $\mu(dt, dx, du)$ on $(0,\infty)^3$, independent of Q and of intensity $e^{-u}\,dt\,dx\,du$, such that, for $t \le T_1$,

$$NX_t^{m+1} = \mu\big(\{(s,x,u) : s \le t < s+u, x \le N\lambda(X_{s-}^m)^d\}\big)$$

and

$$N(X_t^m + X_t^{m+1}) = Q_t + \mu\big(\{(s,x,u) : s \le t < s+u, x \le N\lambda(X_{s-}^{m-1})^d\}\big).$$

Here the u variable encodes the exponential service time of the current customer in each queue.

Define $(\hat{X}_t^m)_{t\ge 0}$ by

$$N\hat{X}_t^m = Q_t + \mu\big(\{(s,x,u) : s \le t < s+u, x \le N\lambda(x_s^{m-1})^d\}\big)$$

and set $T_2 = \inf\{t \ge 0 : X_t^m + X_t^{m+1} \ne \hat{X}_t^m\}$. Then $(N\hat{X}_t^m)_{t\ge 0}$ is an $M/M/\infty$ queue starting from Nx_0^m, with arrival rate $N\lambda(x_t^{m-1})^d$ and service rate 1. Fix $r > 1$ and set $T_3 = \inf\{t \ge 0 : \hat{X}_t^m > ra_m\}$. Fix $R \ge 1$, set

$$T_4 = \inf\big\{t \ge 0 : \sqrt{N}|X_t^k - x_t^k| > R\sqrt{a_k} \text{ for some } k \le m-1\big\},$$

and set $T = T_1 \wedge T_2 \wedge T_3 \wedge T_4 \wedge t_0$. Finally, set

$$p = p(N, \lambda, d, x_0, A, R, r) = \mathbb{P}(T < t_0).$$

PROPOSITION 3.1. *Assume that $x_0 \in S(\rho, t_0)$, that $A, R, \rho \ge 1$ with $\rho A^d \le N^{d-1}$, that $r > \rho$ and that*

$$2rAt_0e^{Lt_0}/N^{(1/2)(1-1/d)} \le R \le (t_0 \wedge 1)\sqrt{A},$$

where $L = 2(d\sigma^{d-1} + 1)$ and $\sigma = \rho + 1$. Then $p \le p_1 + p_2 + p_3 + p_4$, where

$$p_1 = A^d r^d t_0 / N^{d-1},$$

$$p_2 = A^{1-1/(2d)} d\sigma^{d-1} R t_0 / N^{(1/2)(1-1/d)},$$

$$p_3 = \rho^d t_0/(r - \rho),$$

$$p_4 = 2m\exp\big(-R^2/(20\sigma^d t_0 e^{2Lt_0})\big).$$

PROOF. It will suffice to show that $\mathbb{P}(T = T_i) \le p_i$ for $i = 1, 2, 3, 4$. Recall that, for a Poisson random variable Y of parameter $\nu > 0$ and for $a > 0$, we have $\mathbb{P}(Y \ge a) \le \nu/a$. For $t < T$, we have $N\lambda(X_t^m)^d \le N\lambda(ra_m)^d \le A^d r^d/N^{d-1}$, so X_T^{m+1} is dominated by a Poisson random variable Y_1 of parameter p_1, and so

$$\mathbb{P}(T = T_1) = \mathbb{P}(X_T^{m+1} = 1) \le \mathbb{P}(Y_1 \ge 1) \le p_1.$$

Since $x_0 \in S(\rho, t_0)$, we have $x_t^k \le \rho a_k$ for all $k \in \mathbb{N}$ and all $t \le t_0$. For $k \le m-1$, we have $R\sqrt{a_k/N} \le a_k R/\sqrt{Na_k} \le a_k R/\sqrt{A} \le a_k$. Hence, for $t < T$ and $k \le m-1$,

$$X_t^k \le x_t^k + R\sqrt{a_k/N} \le \sigma a_k. \tag{3}$$

Then, for $t < T$,

$$\begin{aligned} N\lambda|(X_t^{m-1})^d - (x_t^{m-1})^d| &\le N\lambda d\sigma^{d-1} a_{m-1}^{d-1} |X_t^{m-1} - x_t^{m-1}| \\ &\le N\lambda d\sigma^{d-1} a_{m-1}^{d-1} R\sqrt{a_{m-1}/N} \\ &\le A^{1-1/(2d)} d\sigma^{d-1} R/N^{(1/2)(1-1/d)}, \end{aligned}$$

where the final inequality follows from $a_m = \lambda a_{m-1}^d$. Set

$$\Delta = \mu\bigl(\{(t, x, u) : t \le T, N\lambda(x_t^{m-1} \wedge X_{t-}^{m-1})^d < x \le N\lambda(x_t^{m-1} \vee X_{t-}^{m-1})^d\}\bigr).$$

Then Δ is dominated by a Poisson random variable Y_2 of parameter p_2. Hence,

$$\mathbb{P}(T = T_2) = \mathbb{P}(\Delta = 1) \le \mathbb{P}(Y_2 \ge 1) \le p_2.$$

Note that $\lambda(x_t^{m-1})^d \le \lambda\rho^d a_{m-1}^d = \rho^d a_m$ for all $t \le t_0$. Thus, $N\hat{X}_T^m \le Nx_0^m + Y_3$ for a Poisson random variable Y_3 of parameter $N\rho^d a_m t_0$ and so

$$\mathbb{P}(T = T_3) = \mathbb{P}(\hat{X}_T^m \ge ra_m) \le \mathbb{P}\bigl(Y_3 \ge N(r - \rho)a_m\bigr) \le p_3.$$

It remains to estimate $\mathbb{P}(T = T_4)$. For this, we write

$$X_t^k = x_0^k + M_t^k + \int_0^t b^k(X_s)\,ds \tag{4}$$

so that

$$X_t^k - x_t^k = M_t^k + \int_0^t \bigl(b^k(X_s) - b^k(x_s)\bigr)\,ds.$$

Then we use a combination of exponential martingale inequalities and Gronwall's lemma to obtain the desired estimate. First we investigate how small M will need to be to obtain the required bound on $|X_t^k - x_t^k|$ for $k \le m-1$. Note that

$$|\lambda_+^k(x) - \lambda_+^k(y)| \le \lambda d(x^{k-1} \vee y^{k-1})^{d-1}|x^{k-1} - y^{k-1}| + \lambda d(x^k \vee y^k)^{d-1}|x^k - y^k|$$

so, provided that $x^k \vee y^k \le \sigma a_k$ and $x^{k-1} \vee y^{k-1} \le \sigma a_{k-1}$,

$$|\lambda_+^k(x) - \lambda_+^k(y)| \le d\sigma^{d-1}\{(a_k/a_{k-1})|x^{k-1} - y^{k-1}| + |x^k - y^k|\}. \tag{5}$$

Also,

$$|\lambda_-^k(x) - \lambda_-^k(y)| \le |x^k - y^k| + |x^{k+1} - y^{k+1}|. \tag{6}$$

Hence, provided that $x^k \vee y^k \le \sigma a_k$ and $x^{k-1} \vee y^{k-1} \le \sigma a_{k-1}$, we have

$$|b^k(x) - b^k(y)|/\sqrt{a_k} \le L \sup_{j=k-1,k,k+1} |x^j - y^j|/\sqrt{a_j}.$$

We note that the definitions of T_2 and T_3 force $X_t^m \le r a_m$ for all $t < T$. Set

$$f(t) = \sup_{k \le m-1} \sup_{s \le t} |X_s^k - x_s^k|/\sqrt{a_k}.$$

Then, for $t < T$ and $k \le m-2$,

$$|b^k(X_t) - b^k(x_t)|/\sqrt{a_k} \le Lf(t)$$

and

$$|b^{m-1}(X_t) - b^{m-1}(x_t)|/\sqrt{a_{m-1}} \le Lf(t) + r a_m/\sqrt{a_{m-1}}.$$

Hence, for $t \le T$,

$$f(t) \le \big(M_t^* + r a_m t/\sqrt{a_{m-1}}\big) + L\int_0^t f(s)\,ds,$$

where

$$M_t^* = \sup_{k \le m-1} \sup_{s \le t} |M_s^k|/\sqrt{a_k}.$$

Set $\alpha_k = \frac{1}{2}e^{-Lt_0}R\sqrt{a_k/N}$ and consider, for $k \le m-1$, the stopping times $T^k = T_-^k \wedge T_+^k$, where

$$T_\pm^k = \inf\{t \ge 0 : \pm M_t^k > \alpha_k\}.$$

Suppose that $T < T^1 \wedge \cdots \wedge T^{m-1}$. Then $M_T^* \le \frac{1}{2}e^{-Lt_0}R/\sqrt{N}$. On the other hand,

$$\begin{aligned} r a_m t_0/\sqrt{a_{m-1}} &= \lambda^{1/(2d)} r a_m^{1-1/(2d)} t_0 \\ &\le r A^{1-1/(2d)} t_0/N^{1-1/(2d)} \le \tfrac{1}{2}e^{-Lt_0}R/\sqrt{N}. \end{aligned}$$

So by Gronwall's lemma,

$$f(T) \le e^{Lt_0}\big(M_T^* + r a_m t_0/\sqrt{a_{m-1}}\big) \le R/\sqrt{N}.$$

Hence,

$$\mathbb{P}(T = T_4) \le \sum_{k=1}^{m-1} \mathbb{P}(T^k \le T)$$

and it remains to estimate $\mathbb{P}(T_{\pm}^k \le T)$ for $k \le m-1$. For $k \in \mathbb{N}$, $x \in S$ and $\theta \in \mathbb{R}$, set

$$\phi^k(x,\theta) = \lambda_+^k(x)h(\theta) + \lambda_-^k(x)h(-\theta),$$

where $h(\theta) = e^\theta - 1 - \theta$. For $t < T$ and $k \le m-1$, we have $X_t^k \le \sigma a_k$ so

$$\phi^k(X_t,\theta) \le \lambda\sigma^d a_{k-1}^d h(\theta) + \sigma a_k h(-\theta) \le \sigma^d a_k g(\theta), \tag{7}$$

where $g(\theta) = e^\theta - 2 + e^{-\theta}$. Consider, for $\theta \ge 0$, the exponential martingale

$$\begin{aligned} Z_t^k &= \exp\left\{N\theta(X_t^k - X_0^k) - \int_0^t \int_{\mathbb{R}^{\mathbb{N}}} (e^{N\theta y^k} - 1) K(X_s, dy)\, ds\right\} \\ &= \exp\left\{N\theta M_t^k - N\int_0^t \phi^k(X_s,\theta)\, ds\right\} \end{aligned}$$

and note that, on the event $T_+^k \le T$, we have

$$Z_{T_+^k}^k \ge \exp\{N\theta\alpha_k - N\sigma^d a_k g(\theta)t_0\}.$$

By optional stopping, $\mathbb{E}(Z_{T_1^k}^{k,N}) \le 1$, so

$$\mathbb{P}(T_+^k \le T) \le \exp(-N\theta\alpha_k + N\sigma^d a_k g(\theta) t_0).$$

We choose $\theta = \alpha_k/(2\sigma^d a_k \gamma t_0)$, where $\gamma = g(1) \le \frac{5}{4}$. Using $R \le t_0\sqrt{A}$ it is straightforward to check that $\theta \le 1$, so from Taylor's theorem, $g(\theta) \le \gamma\theta^2$. Hence,

$$\mathbb{P}(T_+^k \le T) \le \exp(-N\alpha_k^2/(4\sigma^d a_k \gamma t_0)) = \exp(-R^2/(16\sigma^d \gamma t_0 e^{2Lt_0})).$$

The same bound applies to $\mathbb{P}(T_-^k \le T)$. So we have shown that $\mathbb{P}(T_4 = T) \le p_4$, as required. $\square$

PROOF OF THEOREM 2.1. We will determine conditions on sequences $R(N)$ and $r(N)$ so that, for $A(N) = (\log N)^4$, as $N \to \infty$, all the constraints of Proposition 3.1 are satisfied and, with an obvious notation, $p_i(N) \to 0$ for $i = 1,2,3,4$. For $p_4(N) \to 0$, it suffices that $\log\log N \exp(-R^2/(20\sigma^d t_0 e^{2Lt_0})) \to 0$ and, hence, that $R/\sqrt{\log\log\log N} \to \infty$. For $p_3(n) \to 0$, it suffices that $r \to \infty$. For $p_2(N) \to 0$, it suffices that $A^{2-1/d}R^2/N^{1-1/d} \to 0$ and for $p_1(N) \to 0$, it suffices that $Ar/N^{1-1/d} \to 0$. If we can also arrange that $R/\sqrt{A} \to 0$ and $rA/(RN^{(1/2)(1-1/d)}) \to 0$, then all the constraints of Proposition 3.1 will be satisfied eventually. A possible choice is to take $r(N) = N^{(1/2)(1-1/d)}/(\log N)^4$ and any sequence $R(N)$ with $R(N)/\sqrt{\log\log\log N} \to \infty$ and $R(N)/(\log N)^2 \to 0$. This proves the first part of the theorem. For the final assertion, it suffices to note that, for $k \le m-1$, $R(N)/\sqrt{Na_k} \le R(N)/\sqrt{A(N)} \to 0$. $\square$

We remark that the choice $R(N) = \log N$ leads to a bound of the form $p(N) \le CN^{-(1/2)(1-1/d)}$ up to logarithmic corrections. This is the best rate of decay of probabilities we have found. We remark also that a marginally shorter proof can be had by replacing the exponential martingale inequality by Doob's L^2-inequality, at the small cost of requiring that $R(N)/\sqrt{\log\log N} \to \infty$.

4. A refinement of the fluid limit. This section leads to a proof of Theorem 2.2. The deterministic limit (for components $k \le m-1$) just discussed will be refined by approximating the martingale M in (4) by another martingale whose characteristics are determined by the limit path, and at the same time linearizing around the limit path. The accuracy of the approximation is thereby improved from $N^{-1/2}$ to $N^{-3/4}$ at the cost of moving to an approximating process which is not deterministic but has a simple random structure, being a linear function of a Poisson random measure.

Define a measure $\tilde{\nu}$ on $\mathbb{R}^{\mathbb{N}} \times (0, t_0]$ by

$$\tilde{\nu}(dy, dt) = K(x_t, dy)\, dt.$$

We will take $\tilde{\mu}$ to be a Poisson random measure with intensity $\tilde{\nu}$ coupled, in a way to be specified, with the process X. Define $\tilde{M} = (\tilde{M}_t^k : k \in \mathbb{N}, t \le t_0)$ by

$$\tilde{M}_t^k = \int_{\mathbb{R}^{\mathbb{N}} \times (0,t]} y^k (\tilde{\mu} - \tilde{\nu})(dy, ds)$$

and define $\tilde{\gamma} = (\tilde{\gamma}_t^k : k \in \mathbb{N}, t \le t_0)$ by

$$\tilde{\gamma}_t = \sqrt{N}\tilde{M}_t + \int_0^t \nabla b(x_s)\tilde{\gamma}_s\, ds.$$

We show in Section 6 that we can write $\tilde{\gamma}$ as an explicit linear function of $\tilde{\mu} - \tilde{\nu}$,

$$\tilde{\gamma}_t = \sqrt{N} \int_{\mathbb{R}^{\mathbb{N}} \times (0,t]} \Phi_{t,s}\, y(\tilde{\mu} - \tilde{\nu})(dy, ds), \tag{8}$$

where $(\Phi_{t,s} : s \le t \le t_0)$ is the $\mathbb{N} \times \mathbb{N}$ matrix-valued process given by

$$\frac{\partial}{\partial t}\Phi_{t,s} = \nabla b(x_t)\Phi_{t,s}, \qquad \Phi_{s,s} = I.$$

Thus, $\tilde{\gamma}$ has a simpler stochastic structure than X. In particular, we can write the characteristic function of any finite-dimensional distribution of $\tilde{\gamma}$ in terms of $(x_t)_{t \le t_0}$ and $(\Phi_{t,s} : s \le t \le t_0)$.

Recall that

$$X_t = x_0 + M_t + \int_0^t b(X_s)\, ds.$$

On the other hand, if we set $\tilde{X}_t = x_t + N^{-1/2}\tilde{\gamma}_t$, then

$$\tilde{X}_t = x_0 + \tilde{M}_t + \int_0^t b(x_s)\, ds + \int_0^t \nabla b(x_s)(\tilde{X}_s - x_s)\, ds.$$

Set $\tilde{Y} = X - \tilde{X}$, $\tilde{D} = M - \tilde{M}$ and

$$A_t = \int_0^t \big(b(X_s) - b(x_s) - \nabla b(x_s)(X_s - x_s)\big)\, ds. \tag{9}$$

Then

$$\tilde{Y}_t = \tilde{D}_t + A_t + \int_0^t \nabla b(x_s)\tilde{Y}_s\, ds. \tag{10}$$

We will obtain a good approximation if we can couple $\tilde{M}$ with M to make $\tilde{D}$ small. Define kernels K_0, K_+, K_- on $(0, t_0] \times E \times \mathbb{R}^{\mathbb{N}}$ by

$$K_0(t, x, dy) = K(x, dy) \wedge K(x_t, dy),$$
$$K_\pm(t, x, dy) = \big(K(x, dy) - K(x_t, dy)\big)^\pm,$$

and let $K_*(t, x, w, dx', dw')$ be the image of the measure

$$K_0(t, x, dy_0) \otimes K_+(t, x, dy_+) \otimes K_-(t, x, dy_-)$$

by the map $(x', w') = (y_0 + y_+, y_0 + y_-)$. Let $(X_t, W_t)_{t \geq 0}$ be a Markov chain, starting from $(x_0, 0)$, with time-dependent Lévy kernel K_*. Set

$$\tilde{\mu}_t = \sum_{\Delta W_t \neq 0} \delta_{(t, \Delta W_t)}.$$

Then $(X_t)_{t \geq 0}$ is a Markov chain with Lévy kernel K and $\tilde{\mu}$ is a Poisson random measure with intensity $\tilde{\nu}$. We have coupled X and W so that, as far as possible, they have the same jumps. Set $T_5 = \inf\{t \geq 0 : |\tilde{\gamma}_t^m| > r\sqrt{a_m}\}$, where r is as in the previous section. Fix $\tilde{R} > 0$ and set

$$T_6 = \inf\{t \geq 0 : N^{3/4}|X_t^k - \tilde{X}_t^k| > \tilde{R} a_k^{1/4} \text{ for some } k \leq m - 1\}.$$

Finally, set $\tilde{T} = T \wedge T_5 \wedge T_6$ and $\tilde{p} = \mathbb{P}(\tilde{T} < t_0)$.

PROPOSITION 4.1. *Assume that the conditions of Proposition* 3.1 *hold. Set* $H = \frac{1}{2} d(d-1)\sigma^{d-2}$ *and assume, in addition, that* $rA \leq N^{(1/2)(1-1/d)}$ *and*

$$4HR^2 t_0 e^{Lt_0}/N^{1/4} + 4rAt_0 e^{Lt_0}/N^{(1/4)(1-1/d)} \leq \tilde{R} \leq 4RLt_0 e^{Lt_0} A^{1/4}.$$

Then $\tilde{p} \leq p_1 + p_2 + p_3 + p_4 + p_5 + p_6$, *where* p_1, p_2, p_3, p_4 *are defined in Proposition* 3.1 *and*

$$p_5 = 8(\rho^d + 1) t_0 e^{2Lt_0^2}/r^2, \qquad p_6 = 2m \exp\{-\tilde{R}^2/(20RLt_0 e^{2Lt_0})\}.$$

PROOF. Given Proposition 3.1, it will suffice to show that $\mathbb{P}(\tilde{T} = T_i) \leq p_i$ for $i = 5, 6$. By Proposition 6.1,

$$\mathbb{E}\Big(\sup_{t \leq t_0} |\tilde{\gamma}_t^m|^2\Big) \leq 8(\rho^d + 1) t_0 e^{2Lt_0^2} a_m.$$

So $\mathbb{P}(\tilde{T} = T_5) \le p_5$ by Chebyshev's inequality.

We now follow an argument similar to the proof that $\mathbb{P}(T = T_4) \le p_4$ in Proposition 3.1. Set

$$\tilde{f}(t) = \sup_{k \le m-1} \sup_{s \le t} |\tilde{Y}_s^k| / a_k^{1/4},$$

$$\tilde{A}_t^* = \sup_{k \le m-1} \sup_{s \le t} |A_s^k| / a_k^{1/4} + \int_0^t |\tilde{Y}_s^m| \, ds / a_{m-1}^{1/4},$$

$$\tilde{D}_t^* = \sup_{k \le m-1} \sup_{s \le t} |\tilde{D}_s^k| / a_k^{1/4}.$$

We recall that

$$\nabla b^k(x) y = \lambda d (x^{k-1})^{d-1} y^{k-1} - \lambda d (x^k)^{d-1} y^k - y^k + y^{k+1}$$

so, provided that $x^k \le \sigma a_k$ and $x^{k-1} \le \sigma a_{k-1}$,

$$|\nabla b^k(x) y| / a_k^{1/4} \le L \sup_{j = k-1, k, k+1} |y^j| / a_j^{1/4}.$$

For $t \le t_0$, we have $x_t^k \le \rho a_k \le \sigma a_k$ for all k, so, for $k \le m-1$,

$$|\nabla b^k(x_t) \tilde{Y}_t| / a_k^{1/4} \le L \tilde{f}(t) + \delta_{k,m-1} |\tilde{Y}_t^m| / a_{m-1}^{1/4}.$$

Then, from (10), we get

$$\tilde{f}(t) \le \tilde{D}_t^* + \tilde{A}_t^* + L \int_0^t \tilde{f}(s) \, ds,$$

so, by Gronwall's lemma, $\tilde{f}(t) \le e^{Lt}(\tilde{A}_t^* + \tilde{D}_t^*)$ for all $t \le t_0$. Note that, for $k \le m-1$,

$$\begin{aligned} b^k(y) &- b^k(x) - \nabla b^k(x)(y - x) \\ &= \lambda\big((y^{k-1})^d - (x^{k-1})^d - d(x^{k-1})^{d-1}(y^{k-1} - x^{k-1})\big) \\ &\quad - \lambda\big((y^k)^d - (x^k)^d - d(x^k)^{d-1}(y^k - x^k)\big), \end{aligned}$$

so, provided that $x^k, y^k \le \sigma a_k$ and $x^{k-1}, y^{k-1} \le \sigma a_{k-1}$,

$$\begin{aligned} &|b^k(y) - b^k(x) - \nabla b^k(x)(y - x)| \\ &\quad \le H\lambda(a_{k-1}^{d-2} |y^{k-1} - x^{k-1}|^2 + a_k^{d-2} |y^k - x^k|^2). \end{aligned}$$

For $t < \tilde{T}$ and $k \le m-1$, we have $|X_t^k - x_t^k| \le R\sqrt{a_k/N}$; moreover, as we showed at (3), this implies that $X_t^k \le \sigma a_k$. Hence,

$$\begin{aligned} &|b^k(X_t) - b^k(x_t) - \nabla b^k(x_t)(X_t - x_t)| / a_k^{1/4} \\ &\quad \le H\lambda R^2 (a_{k-1}^{d-1} + a_k^{d-1}) / (N a_k^{1/4}) \le 2CR^2/N. \end{aligned}$$

Also, $|\tilde{Y}_t^m| \le |X_t^m - x_t^m| + N^{-1/2}|\tilde{\gamma}_t^m|$, so, for $t < \tilde{T}$, $|\tilde{Y}_t^m|/a_{m-1}^{1/4} \le (ra_m + N^{-1/2}r\sqrt{a_m})/a_{m-1}^{1/4} \le 2r(A/N)^{1-1/(4d)}$. It follows that

$$\tilde{A}^*_{\tilde{T}} \le 2HR^2 t_0/N + 2rt_0(A/N)^{1-1/(4d)} \le \tfrac{1}{2}e^{-Lt_0}\tilde{R}/N^{3/4}.$$

Set $\tilde{\alpha}_k = \frac{1}{2}e^{-Lt_0}\tilde{R}a_k^{1/4}/N^{3/4}$ and consider the stopping times $\tilde{T}^k = \tilde{T}^k_+ \wedge \tilde{T}^k_-$, where

$$\tilde{T}^k_\pm = \inf\{t \ge 0 : \pm\tilde{D}^k_t > \tilde{\alpha}_k\}.$$

Suppose that $\tilde{T} < \tilde{T}^1 \wedge \cdots \wedge \tilde{T}^{m-1}$. Then $\tilde{D}^*_{\tilde{T}} \le \frac{1}{2}e^{-Lt_0}\tilde{R}/N^{3/4}$ so $\tilde{f}(\tilde{T}) \le \tilde{R}/N^{3/4}$ and $\tilde{T} < T_6$. Hence,

$$\mathbb{P}(\tilde{T} = T_6) \le \sum_{k=1}^{m-1} \mathbb{P}(\tilde{T}^k \le \tilde{T})$$

and it remains to estimate $\mathbb{P}(\tilde{T}^k_\pm \le \tilde{T})$ for $k \le m-1$.

For $k \le m-1$, set

$$\psi^k(t,x,\theta) = \big(\lambda^k_+(x) - \lambda^k_+(x_t)\big)^+ h(\theta) + \big(\lambda^k_-(x) - \lambda^k_-(x_t)\big)^+ h(-\theta) + \big(\lambda^k_+(x) - \lambda^k_+(x_t)\big)^- h(-\theta) + \big(\lambda^k_-(x) - \lambda^k_-(x_t)\big)^- h(\theta).$$

Fix $\theta \ge 0$ and consider, for $k \le m-1$, the exponential martingale

$$\begin{aligned}\tilde{Z}^k_t &= \exp\bigg\{N\theta(X^k_t - X^k_0 - W^k_t) - \int_0^t \int_{\mathbb{R}^{\mathbb{N}}} (e^{N\theta y^k} - 1)K^+(s, X_s, dy)\,ds \\ &\qquad - \int_0^t \int_{\mathbb{R}^{\mathbb{N}}} (e^{-N\theta y^k} - 1)K^-(s, X_s, dy)\,ds\bigg\} \\ &= \exp\bigg\{N\theta\tilde{D}^k_t - N\int_0^t \psi^k(s, X_s, \theta)\,ds\bigg\}.\end{aligned}$$

For $k \le m-2$ and $t \le t_0$, for $x^k \le \sigma a_k$ and $x^{k-1} \le \sigma a_{k-1}$, we can estimate as at (5), (6) and (7) to obtain

$$\begin{aligned}\psi^k(t,x,\theta)/\sqrt{a_k} &\le g(\theta)\big(|\lambda^k_+(x) - \lambda^k_+(x_t)| + |\lambda^k_-(x) - \lambda^k_-(x_t)|\big)/\sqrt{a_k} \\ &\le Lg(\theta) \sup_{j=k-1,k,k+1} |x^j - x^j_t|/\sqrt{a_j},\end{aligned}$$

so, for $t < \tilde{T}$ and $k \le m-2$,

$$\psi^k(t, X_t, \theta) \le Lg(\theta)R\sqrt{a_k/N}. \tag{11}$$

Similarly, since we assume $R \ge 1, rA \le N^{(1/2)(1-1/d)}$ and $Na_m \le A$, we have, for $t < \tilde{T}$, $X^m_t \le ra_m \le R\sqrt{a_{m-1}/N}$ and we can show that (11) remains true for

$k = m - 1$. By optional stopping we have $\mathbb{E}(\tilde{Z}^k_{\tilde{T}^k_+}) \le 1$ for all $k \le m-1$. But on the event $\tilde{T}^k_+ \le \tilde{T}$, we have

$$\tilde{Z}^k_{\tilde{T}^k_+} \ge \exp\{N\theta\tilde{\alpha}^k - Lg(\theta)Rt_0\sqrt{Na_k}\}.$$

We choose $\theta = 2\sqrt{N}\tilde{\alpha}^k/(5LRt_0\sqrt{a_k})$, checking that $\theta \le 1$, so that $g(\theta) \le 5\theta^2/4$, and deduce

$$\begin{aligned}\mathbb{P}(\tilde{T}^k_+ \le \tilde{T}) &\le \exp\{-N^{3/2}\tilde{\alpha}_k^2/(5LRt_0\sqrt{a_k})\}\\ &= \exp\{-\tilde{R}^2/(20LRt_0e^{2Lt_0})\}.\end{aligned}$$

The same bound applies to $\mathbb{P}(\tilde{T}^k_- \le \tilde{T})$. So we have shown that $\mathbb{P}(\tilde{T} = T_6) \le p_6$, as required. □

PROOF OF THEOREM 2.2. Choose $r(N)$ as in the proof of Theorem 2.1, so that $r(N) \to \infty$ and so $p_5(N) \to 0$. Assume $\tilde{R}(N)/(\log N)^2 \to 0$, and set $s(N) = \tilde{R}(N)/(\log\log\log N)^{3/4}$. Then $s(N) \to \infty$. Set $R(N) = s(N)(\log\log\log N)^{1/2}$ and $r(N) = N^{(1/4)(1-1/d)}/(\log N)^4$. It is straightforward to check that all the constraints in Propositions 3.1 and 4.1 are satisfied eventually. Moreover, as in the proof of Theorem 2.1, we have $p_i(N) \to 0$ for $i = 1, 2, 3, 4$. Finally, $\tilde{R}(N)^2/R\log\log\log N \to \infty$, so also $p_6(N) \to 0$, which proves the theorem. □

5. Diffusion approximation. In this section we prove Theorem 2.3. The method follows the lines set out in [4], Chapter 11. As we have already seen, our process X has around $\log\log N$ active components, which have a wide range of scales. This will require special consideration in the implementation of the general method. We also have to deal with the fact that the variance of the diffusion approximation has degeneracies. The diffusion coefficient, obtained as the square root of the variance, then fails to be Lipschitz and some special care is needed to arrive at the desired convergence.

Let $(X^k_t : k \in \mathbb{N}, t \ge 0)$ be the supermarket process starting from x_0 and recall equation (4)

$$X^k_t = x^k_0 + M^k_t + \int_0^t b^k(X_s)\,ds.$$

Recall also that we set $\bar{X}_t = x_t + N^{-1/2}\gamma_t$, where $(\gamma^k_t : k \in \mathbb{N}, t \le t_0)$ is defined by the linear equations (2)

$$\gamma^k_t = \sqrt{N}\bar{M}^k_t + \int_0^t \nabla b^k(x_s)\gamma_s\,ds$$

and

$$\sqrt{N}\bar{M}^k_t = \int_0^t \sigma^k_+(x_s)\,dB^k_+(s) - \int_0^t \sigma^k_-(x_s)\,dB^k_-(s).$$

Set $Y = X - \bar{X}$ and $D = M - \bar{M}$. Then

$$Y_t = D_t + A_t + \int_0^t \nabla b^k(x_s) Y_s\, ds,$$

where A_t is defined at (9). We will obtain a good approximation if we can couple $\bar{M}$ with M to make D small.

The coupling relies on the following approximation result of [8]: there exists a constant $c \in (0, \infty)$ and a probability space on which are defined a compensated Poisson process Z of rate 1 and a standard Brownian motion W such that, for all $t \geq 0$ and $x \in \mathbb{R}$,

$$\mathbb{P}\Bigl(\sup_{s \leq t} |Z(s) - W(s)| \geq c \log t + x\Bigr) \leq c e^{-x/c}. \tag{12}$$

See [13] for a recent review of developments and clarifications in connection with this result.

Given independent compensated Poisson processes $Z_+^k, Z_-^k, k \in \mathbb{N}$ of rate 1, we can construct X by the equations (4) and

$$M_t^k = N^{-1}\biggl\{Z_+^k\biggl(N \int_0^t \lambda_+^k(X_s)\, ds\biggr) - Z_-^k\biggl(N \int_0^t \lambda_-^k(X_s)\, ds\biggr)\biggr\}.$$

On the other hand, by a theorem of Knight, see, for example, [15], there exist independent Brownian motions $W_+^k, W_-^k, k \in \mathbb{N}$, such that, for all $k \in \mathbb{N}$ and $t \leq t_0$,

$$W_\pm^k\biggl(N \int_0^t \lambda_\pm^k(\bar{X}_s)\, ds\biggr) = \sqrt{N} \int_0^t \sigma_\pm^k(\bar{X}_s)\, dB_\pm^k(s).$$

The law of $(B_+^k, B_-^k : k \in \mathbb{N})$, given $(W_+^k, W_-^k : k \in \mathbb{N})$, is given by a measurable kernel. So we may assume that these processes are defined on the same probability space as $(Z_+^k, Z_-^k : k \in \mathbb{N})$ and that (Z_+^k, W_+^k), (Z_-^k, W_-^k), $k \in \mathbb{N}$, are independent copies of (Z, W).

Set $T_7 = \inf\{t \geq 0 : |\gamma_t^m| > r\sqrt{a_m}\}$. Fix $\bar{R} > 0$ and set

$$T_8 = \inf\{t \geq 0 : N|X_t^k - \bar{X}_t^k| > \bar{R} \log(N a_k) \text{ for some } k \leq m - 1\}.$$

Finally, set $\bar{T} = T \wedge T_7 \wedge T_8$ and $\bar{p} = \mathbb{P}(\bar{T} < t_0)$.

PROPOSITION 5.1. *Assume that the conditions of Proposition* 3.1 *hold, and assume, moreover, that $A \geq e^2$ and $R \leq \sqrt{A}/2$. There is a constant $C < \infty$, depending only on d, λ, ρ and t_0, such that, if*

$$C + C(R^2 + rA)/\log N \leq \bar{R} \leq A/(2 \log A),$$

then $\bar{p} \leq p_1 + p_2 + p_3 + p_4 + p_7 + p_8$, where p_1, p_2, p_3, p_4 are defined in Proposition 3.1 *and*

$$p_7 = C/r^2, \qquad p_8 = Cm\bigl(A^{-1} + (\bar{R} \log A)^{-2}\bigr).$$

PROOF. Given Proposition 3.1, it will suffice to show that $\mathbb{P}(\bar{T} = T_i) \le p_i$ for $i = 7, 8$. By Proposition 6.1,

$$\mathbb{E}\Big(\sup_{t \le t_0} |\gamma_t^m|^2\Big) \le 8(\rho^d + 1) t_0 e^{2Lt_0^2} a_m.$$

So $\mathbb{P}(\bar{T} = T_7) \le p_7$ for a suitably large C by Chebyshev's inequality. Set

$$\bar{f}(t) = \sup_{k \le m-1} \sup_{s \le t} |Y_s^k| / \log(N a_k),$$

$$A_t^* = \sup_{k \le m-1} \sup_{s \le t} |A_s^k| / \log(N a_k) + \int_0^t |Y_s^m| \, ds / \log(N a_{m-1}),$$

$$D_t^* = \sup_{k \le m-1} \sup_{s \le t} |D_s^k| / \log(N a_k).$$

Since $N a_{m-1} > A \ge e$, we have $N a_k / \log(N a_k) \le N a_{k-1} / \log(N a_{k-1})$ for all $k \le m-1$. So we can use an argument from the proof of Proposition 4.1 to obtain $\bar{f}(t) \le e^{Lt}(A_t^* + D_t^*)$ for all $t \le t_0$.

The function $F(s) = \log s / (N^{1-d} \lambda s^d)$ is decreasing when $s > e$ and $e < A < N a_{m-1} \le (A N^{d-1}/\lambda)^{1/d}$, so

$$\log(N a_{m-1})/(N a_m) = F(N a_{m-1}) \ge F\big((A N^{d-1}/\lambda)^{1/d}\big) \ge \log N / (2A).$$

Similarly, $\log(N a_{m-1})/\sqrt{N a_m} \ge \log N / (2\sqrt{A})$. Hence,

$$r\big(N a_m + \sqrt{N a_m}\big)/\log(N a_{m-1}) \le 4 r A / \log N.$$

We estimate as in the proof of Proposition 4.1 to obtain, for $t < \bar{T}$ and $k \le m-1$,

$$\begin{aligned} &|b^k(X_t) - b^k(x_t) - \nabla b^k(x_t)(X_t - x_t)| / \log(N a_k) \\ &\quad \le H \lambda R^2 (a_{k-1}^{d-1} + a_k^{d-1}) / N \log(N a_k) \le 2 H R^2 / (N \log N) \end{aligned}$$

and

$$|Y_t^m| / \log(N a_{m-1}) \le \big(r a_m + N^{-1/2} r \sqrt{a_m}\big) / \log(N a_{m-1}) \le 4 r A / (N \log N).$$

So $A_{\bar{T}}^* \le 2 t_0 (H R^2 + 2 r A)/(N \log N) \le \frac{1}{2} e^{-L t_0} \bar{R}/N$, provided C is chosen suitably large. For $k \le m-1$, set $\alpha_k = \frac{1}{2} e^{-L t_0} \bar{R} \log(N a_k)/N$ and consider the event $\Omega_k = \{\sup_{t \le \bar{T}} |D_t^k| > \alpha_k\}$. On $\Omega_0 = \Omega \setminus (\Omega_1 \cup \cdots \cup \Omega_{m-1})$, we have $D_{\bar{T}}^* \le \frac{1}{2} e^{-L t_0} \bar{R}/N$, so $\bar{f}(\bar{T}) \le \bar{R}/N$ and $\bar{T} < T_8$. Hence,

$$\mathbb{P}(\bar{T} = T_8) \le \sum_{k=1}^{m-1} \mathbb{P}(\Omega_k)$$

and it will suffice to estimate $\mathbb{P}(\Omega_k)$ for each $k \le m-1$. Fix $k \le m-1$. We can write $D_t^k = D_+(t) - D_-(t)$, where $D_\pm(t) = D_\pm^1(t) + D_\pm^2(t) + D_\pm^3(t)$ and

$$D_\pm^1(t) = N^{-1}(Z_\pm^k - W_\pm^k)\left(N\int_0^t \lambda_\pm^k(X_s)\,ds\right),$$

$$D_\pm^2(t) = N^{-1}\left\{W_\pm^k\left(N\int_0^t \lambda_\pm^k(X_s)\,ds\right) - W_\pm^k\left(N\int_0^t \lambda_\pm^k(\bar{X}_s)\,ds\right)\right\},$$

$$D_\pm^3(t) = N^{-1/2}\int_0^t \{\sigma_\pm^k(\bar{X}_s) - \sigma_\pm^k(x_s)\}\,dB_\pm^k(s).$$

Hence, $\mathbb{P}(\Omega_k) \le q_+^1 + q_-^1 + q_+^2 + q_-^2 + q_+^3 + q_-^3$, where, for $j = 1,2,3$, $q_\pm^j = \mathbb{P}(\Omega_\pm^j)$ and $\Omega_\pm^j = \{\sup_{t\le\bar{T}} |D_\pm^j(t)| > \alpha_k/6\}$.

For $t < \bar{T}$, we have $\lambda_\pm^k(X_t) \le \sigma^d a_k$, so, taking $t = N\sigma^d a_k t_0$ and $x = N\alpha_k/6 - c\log t$ in (12), we obtain

$$\begin{aligned} q_\pm^1 &\le \mathbb{P}\left(\sup_{t\le N\sigma^d a_k t_0} |Z(t) - W(t)| > N\alpha_k/6\right) \\ &\le c\sigma^d t_0 N a_k e^{-N\alpha_k/(6c)} = c\sigma^d t_0 N a_k e^{-\bar{R}\log(Na_k)/(12ce^{Lt_0})} \\ &= c\sigma^d t_0 (Na_k)^{1-\bar{R}/(12ce^{Lt_0})} \le c\sigma^d t_0 A^{1-\bar{R}/(12ce^{Lt_0})} \le C/(4A), \end{aligned}$$

for a suitable choice of C.

We turn to estimate $q_\pm^2$. This will rely on the following continuity estimate for Brownian motion: for $\tau, h, \delta > 0$, setting $n = \lfloor \tau/h \rfloor$,

$$\begin{aligned} &\mathbb{P}\left(\sup_{s,t\le\tau,|s-t|\le h} |W(t) - W(s)| > \delta\right) \\ &\quad \le \mathbb{P}\left(\sup_{k\in\{0,1,\dots,n-1\},t\le 2h} |W(kh+t) - W(kh)| > \delta/2\right) \\ &\quad \le 2n\mathbb{P}\left(\sup_{t\le 2h} W(t) > \delta/2\right) \le (2\tau/h)e^{-\delta^2/(16h)}. \end{aligned}$$

For $t < \bar{T}$, we have

$$\begin{aligned} \bar{X}_t^k &\le x_t^k + |X_t^k - x_t^k| + |\bar{X}_t^k - X_t^k| \le \rho a_k + R\sqrt{a_k/N} + \bar{R}\log(Na_k)/N \\ &\le (\rho + R/\sqrt{A} + \bar{R}\log A/A)a_k \le \sigma a_k, \end{aligned}$$

so

$$N\int_0^t \lambda_\pm^k(\bar{X}_s)\,ds \le \sigma^d t_0 N a_k.$$

Also, using (5) and (6), for $t < \bar{T}$,

$$\begin{aligned}&\left|N\int_0^t \big(\lambda_+^k(X_s) - \lambda_+^k(\bar{X}_s)\big)\,ds\right|\\&\quad\le Nd\sigma^{d-1}t_0\bar{R}a_k\{\log(Na_{k-1})/(Na_{k-1}) + \log(Na_k)/(Na_k)\}\\&\quad\le 2d\sigma^{d-1}t_0\bar{R}\log(Na_k)\end{aligned}$$

and

$$\begin{aligned}&\left|N\int_0^t \big(\lambda_-^k(X_s) - \lambda_-^k(\bar{X}_s)\big)\,ds\right|\\&\quad\le \bar{R}t_0\{\log(Na_k) + \log(Na_{k+1})\mathbb{1}_{k\le m-2}\} + rt_0\big(Na_m + \sqrt{Na_m}\big)\delta_{k,m-1}\\&\quad\le 2t_0\bar{R}\log(Na_k),\end{aligned}$$

provided that C is sufficiently large. We take $\tau = \sigma^d t_0 N a_k$, $h = 2d\sigma^{d-1}t_0 \times \bar{R}\log(Na_k)$ and

$$\delta = N\alpha_k/6 = \bar{R}\log(Na_k)/(12e^{Lt_0})$$

to obtain

$$q_\pm^2 \le \frac{\sigma N a_k}{d\bar{R}\log(Na_k)}\exp\left\{-\frac{\bar{R}\log(Na_k)}{4608 d\sigma^{d-1}t_0 e^{2Lt_0}}\right\} \le C/(4A)$$

for a suitable choice of C.

It remains to estimate $q_\pm^3$. We shall show below that there exists a constant C_0 such that, for $t \le t_0$ and all $k \in \mathbb{N}$,

$$\mathbb{E}(|\gamma_t^k|^2) \le C_0(x_t^{k-1} - x_t^k) \wedge (x_t^k - x_t^{k+1}).$$

Then, for $t < \bar{T}$,

$$\begin{aligned}\sqrt{N}|\sigma_-^k(\bar{X}_t) - \sigma_-^k(x_t)| &= \sqrt{N}\left|\sqrt{\big((\bar{X}_t^k)^+ - (\bar{X}_t^{k+1})^+\big)^+} - \sqrt{x_t^k - x_t^{k+1}}\right|\\&= \frac{\sqrt{N}|((\bar{X}_t^k)^+ - (\bar{X}_t^{k+1})^+)^+ - (x_t^k - x_t^{k+1})|}{\sqrt{((\bar{X}_t^k)^+ - (\bar{X}_t^{k+1})^+)^+} + \sqrt{x_t^k - x_t^{k+1}}}\\&\le (|\gamma_t^k| + |\gamma_t^{k+1}|)/\sqrt{x_t^k - x_t^{k+1}},\end{aligned}$$

so, by Doob's L^2-inequality,

$$\mathbb{E}\left(\sup_{t\le\bar{T}}|D_-^3(t)|^2\right) \le 4\mathbb{E}\int_0^{t_0} N^{-1}|\sigma_-^k(\bar{X}_s) - \sigma_-^k(x_s)|^2\,ds \le 16t_0C_0/N^2.$$

Hence,

$$\begin{aligned}q_-^3 &= \mathbb{P}\left(\sup_{t\le\bar{T}}|D_-^3(t)| > \alpha_k/6\right)\\&\le 2304t_0e^{2Lt_0}C_0/\big(\bar{R}\log(Na_k)\big)^2 \le C(\bar{R}\log A)^{-2}/2\end{aligned}$$

for a suitable choice of C.

The argument for q_+^3 is similar. For $t < \bar{T}$,

$$\begin{aligned}
&\sqrt{N}|\sigma_+^k(\bar{X}_t) - \sigma_+^k(x_t)| \\
&\quad = \sqrt{N}\Big|\sqrt{\big(\lambda\big(((\bar{X}_t^{k-1})^+)^d - \lambda\big((\bar{X}_t^k)^+\big)^d\big)^+} - \sqrt{\lambda(x_t^{k-1})^d - \lambda(x_t^k)^d}\Big| \\
&\quad \le \lambda d\{(x_t^{k-1} \vee \bar{X}_t^{k-1})^{d-1}|\gamma_t^{k-1}| + (x_t^k \vee \bar{X}_t^k)^{d-1}|\gamma_t^k|\}/\sqrt{\lambda(x_t^{k-1})^d - \lambda(x_t^k)^d} \\
&\quad \le \lambda d 2^{d-1} \\
&\qquad \times \big((x_t^{k-1})^{d-1} + (N^{-1/2}|\gamma_t^{k-1}|)^{d-1}\big)|\gamma_t^{k-1}|/\sqrt{\lambda(x_t^{k-1} - x_t^k)(x_t^{k-1})^{d-1}} \\
&\qquad + \lambda d 2^{d-1}\big((x_t^k)^{d-1} + (N^{-1/2}|\gamma_t^k|)^{d-1}\big)|\gamma_t^k|/\sqrt{\lambda(x_t^{k-1} - x_t^k)(x_t^{k-1})^{d-1}},
\end{aligned}$$

so, by Doob's L^2-inequality,

$$\begin{aligned}
\mathbb{E}\Big(\sup_{t\le\bar{T}}|D_+^3(t)|^2\Big) &\le 4\mathbb{E}\int_0^{t_0} N^{-1}|\sigma_+^k(\bar{X}_s) - \sigma_+^k(x_s)|^2\,ds \\
&\le 8d^2 2^{2(d-1)} t_0\big(1 + N^{(d-1)/2}C(d)\big)C_0/N^2,
\end{aligned}$$

where $C(d) = \mathbb{E}(W(1)^{2d})$. Hence,

$$\begin{aligned}
q_+^3 &= \mathbb{P}\Big(\sup_{t\le\bar{T}}|D_+^3(t)| > \alpha_k/6\Big) \\
&\le 2152 d^2 2^{2(d-1)} t_0\big(1 + N^{(d-1)/2}C(d)\big)C_0 e^{2Lt_0}/\big(\bar{R}\log(Na_k)\big)^2 \\
&\le C(\bar{R}\log A)^{-2}/2
\end{aligned}$$

for a suitable choice of C. On combining this with the bounds for $q_\pm^1$ and $q_\pm^2$ already found, we obtain the desired bound for p_8. □

PROOF OF THEOREM 2.3. Set $A(N) = r(N) = (\log N)^{1/2}$ and define $\bar{m}(N) = \inf\{k \in \mathbb{N} : Na_k \le A(N)\}$. Set $R(N) = (\log N)^{1/4}(1 \wedge t_0)/2$. It is straightforward to check that, if C is the constant appearing in Proposition 5.1 and if $\bar{R} = 3C$, then all the constraints in Propositions 3.1 and 5.1 are satisfied eventually and, moreover, that $p_i(N) \to 0$ for $i = 1, 2, 3, 4, 7, 8$. Since $\bar{m}(N) \ge m(N)$, this proves the theorem. □

6. Fluctuation variance estimates. We have deferred from other sections the analysis of certain linear equations associated with our processes. The basic questions of existence and uniqueness in suitable spaces may be resolved by standard methods, so we review this only briefly. The more delicate result, Proposition 6.1, which is needed for the diffusion approximation, relies on the particular structure of our model.

We recall the $\mathbb{N} \times \mathbb{N}$ matrix-valued equation

$$\frac{\partial}{\partial t}\Phi_{t,s} = \nabla b(x_t)\Phi_{t,s}, \qquad \Phi_{s,s} = I,$$

to be solved for $0 \le s \le t \le t_0$. Note that, for $x_0 \in S(\rho, t_0)$ and $t \in [0, t_0]$, we have $\|\nabla b(x_t)\| \le L$, where $\| \cdot \|$ is the operator norm corresponding to $\|x\| = \sup_k |x^k|/a_k$. Hence, it is standard that this equation has a unique continuous solution with $\|\Phi_{t,s}\| \le e^{L(t-s)}$ for all s, t.

The other relevant equations may be considered as stochastic perturbations of the preceding equation. In Theorem 2.2 we used (1)

$$\tilde{\gamma}_t^k = \sqrt{N}\tilde{M}_t^k + \int_0^t \nabla b^k(x_s)\tilde{\gamma}_s\,ds, \qquad t \le t_0,$$

and in Theorem 2.3 we used (2)

$$\gamma_t^k = \sqrt{N}\bar{M}_t^k + \int_0^t \nabla b^k(x_s)\gamma_s\,ds, \qquad t \le t_0. \tag{13}$$

Here

$$\tilde{M}_t^k = \int_{\mathbb{R}^{\mathbb{N}} \times (0,t]} y^k(\tilde{\mu} - \tilde{\nu})(dy, ds)$$

and

$$\sqrt{N}\bar{M}_t^k = \int_0^t \sigma_+^k(x_s)\,dB_+^k(s) - \int_0^t \sigma_-^k(x_s)\,dB_-^k(s).$$

Note that

$$\begin{aligned} N\mathbb{E}(|\tilde{M}_t^k|^2) &= N\int_0^t \int_{\mathbb{R}^{\mathbb{N}}} (y^k)^2 K(x_s, dy)\,ds \\ &= \int_0^t \big(\lambda_+^k(x_s) + \lambda_-^k(x_s)\big)\,ds = N\mathbb{E}(|M_t^k|^2) \end{aligned}$$

and $\lambda_+^k(x_t) + \lambda_-^k(x_t) \le (\rho^d + 1)a_k$ for $t \le t_0$. Hence, a standard type of iteration argument, using Doob's L^2-inequality, shows that these equations have unique measurable solutions with, respectively,

$$\mathbb{E}\Big(\sup_{t \le t_0} |\tilde{\gamma}_t^k|^2\Big) \le 8(\rho^d + 1)t_0 e^{2Lt_0^2} a_k$$

and

$$\mathbb{E}\Big(\sup_{t \le t_0} |\gamma_t^k|^2\Big) \le 8(\rho^d + 1)t_0 e^{2Lt_0^2} a_k.$$

The details for (1) follow below; (2) may be treated in the same way. Note that, for any vector y and for all $k \in \mathbb{N}$,

$$\frac{|\nabla b^k(x_s)y^k|}{\sqrt{a_k}} \le L \sum_{j=k-1,k,k+1} \frac{|y_j|}{\sqrt{a_j}}.$$

Now let

$$\tilde{\gamma}_t^{(0)} = \sqrt{N}\tilde{M}_t,$$

and for $n \in \mathbb{N}$,

$$\tilde{\gamma}_t^{(n)} = \sqrt{N}\tilde{M}_t + \int_0^t \nabla b^k(x_s)\tilde{\gamma}_s^{(n-1)}\,ds.$$

Then for each k,

$$\begin{aligned}\frac{|\tilde{\gamma}_t^{(n+1),k} - \tilde{\gamma}_t^{(n),k}|^2}{a_k} &\le L^2\biggl(\int_0^t \sum_{j=k-1,k,k+1} \frac{|\tilde{\gamma}_s^{(n),j} - \tilde{\gamma}_s^{(n-1),j}|}{\sqrt{a_j}}\,ds\biggr)^2 \\ &\le 3L^2 \sum_{j=k-1,k,k+1}\biggl(\int_0^t \frac{|\tilde{\gamma}_s^{(n),j} - \tilde{\gamma}_s^{(n-1),j}|}{\sqrt{a_j}}\,ds\biggr)^2,\end{aligned}$$

so, using Cauchy–Schwarz,

$$\frac{|\tilde{\gamma}_t^{(n+1),k} - \tilde{\gamma}_t^{(n),k}|^2}{a_k} \le 3L^2 t \sum_{j=k-1,k,k+1} \int_0^t \frac{|\tilde{\gamma}_s^{(n),j} - \tilde{\gamma}_s^{(n-1),j}|^2}{a_j}\,ds.$$

Then for all $0 \le s \le t$,

$$\sup_{s\le t} \frac{|\tilde{\gamma}_s^{(n+1),k} - \tilde{\gamma}_s^{(n),k}|^2}{a_k} \le 3L^2 t \sum_{j=k-1,k,k+1} \int_0^t \sup_{u\le s} \frac{|\tilde{\gamma}_u^{(n),j} - \tilde{\gamma}_u^{(n-1),j}|^2}{a_j}\,ds.$$

Let $h^{(n)}(t) = \sup_k \mathbb{E}(\sup_{s\le t} |\tilde{\gamma}_s^{(n+1),k} - \tilde{\gamma}_s^{(n),k}|^2/a_k)$; then for $t \le t_0$,

$$h^{(n)}(t) \le 9L^2 t_0 \int_0^t h^{(n-1)}(s)\,ds.$$

Hence,

$$h^{(n)}(t) \le \frac{2^n n!(3Lt_0)^{2n}}{(2n)!}\tilde{M}(t_0),$$

where

$$\begin{aligned}\tilde{M}(t) &= \sup_k \mathbb{E}\Bigl(\sup_{s\le t} |\tilde{M}_s^k|^2/a_k\Bigr) \\ &\le 4\sup_k \sup_{s\le t} \mathbb{E}(|\tilde{M}_s^k|^2/a_k) \le 4t(\rho^d + 1).\end{aligned}$$

We deduce that $\tilde{\gamma}_t^{(n)}$ converges to a process $\tilde{\gamma}_t$ uniformly on $[0, t_0]$, and that $\tilde{\gamma}_t$ satisfies (1). The uniqueness part of the proof is similar. Let

$$g(t) = \sup_k \mathbb{E}\Bigl(\sup_{s\le t} |\tilde{\gamma}_s^k|^2\Bigr)\Big/ a_k.$$

Then it follows from the above estimates that, for $t \le t_0$,

$$g(t) \le \tilde{M}(t_0) e^{9L^2 t_0^2}.$$

It may be verified by substitution that (8) gives an explicit representation of the solution of (1).

PROPOSITION 6.1. *The solution* $(\gamma_t^k : k \in \mathbb{N}, t \le t_0)$ *to* (2) *satisfies*

$$\sup_{k \in \mathbb{N}} \sup_{t \le t_0} \mathbb{E}(|\gamma_t^k|^2) / \min\{x_t^{k-1} - x_t^k, x_t^k - x_t^{k+1}\} < \infty.$$

PROOF. Note that $\lambda_+^k(x_t) \le d(x_t^{k-1} - x_t^k)$, so

$$\begin{aligned} \dot{x}_t^k - \dot{x}_t^{k+1} &= \lambda_+^k(x_t) - \lambda_+^{k+1}(x_t) - (x_t^k - x_t^{k+1}) + (x_t^{k+1} - x_t^{k+2}) \\ &\ge \lambda_+^k(x_t) - (d+1)(x_t^k - x_t^{k+1}) + (x_t^{k+1} - x_t^{k+2}). \end{aligned}$$

Hence,

$$\int_0^t \lambda_+^k(x_s)\,ds + \int_0^t (x_s^{k+1} - x_s^{k+2})\,ds \le e^{(d+1)t}(x_t^k - x_t^{k+1})$$

and

$$\int_0^t (x_s^k - x_s^{k+1})\,ds \le t e^{(d+1)t}(x_t^k - x_t^{k+1}).$$

Also,

$$\int_0^t \lambda_+^k(x_s)\,ds \le d \int_0^t (x_s^{k-1} - x_s^k)\,ds \le d t e^{(d+1)t}(x_t^{k-1} - x_t^k).$$

Fix $\varepsilon > 0$ and set

$$f(t) = \sup_{k \in \mathbb{N}} \sup_{s \le t} \mathbb{E}(|\gamma_s^k|^2)/\delta_s^k,$$

where $\delta_t^k = \min\{x_t^{k-1} - x_t^k, x_t^k - x_t^{k+1}\} + \varepsilon$. Then $f(t_0) < \infty$. Note that

$$\begin{aligned} &\mathbb{E}\left(\left|\int_0^t \sigma_+^k(x_s)\,dB_+^k(s) - \int_0^t \sigma_-^k(x_s)\,dB_-^k(s)\right|^2\right) \\ &\quad = \int_0^t \lambda_+^k(x_s)\,ds + \int_0^t (x_s^k - x_s^{k+1})\,ds \le (dt+1)e^{(d+1)t}\delta_t^k. \end{aligned}$$

We have

$$\nabla b^k(x)\gamma = \lambda d (x^{k-1})^{d-1}\gamma^{k-1} - \big(\lambda d (x^k)^{d-1} + 1\big)\gamma^k + \gamma^{k+1}.$$

We will make use of the following estimates:

$$
\begin{aligned}
&\mathbb{E}\left(\left|\int_0^t \lambda d(x_s^{k-1})^{d-1}\gamma_s^{k-1}\,ds\right|^2\right)\\
&\quad\le \int_0^t \lambda^2 d^2 (x_s^{k-1})^{2(d-1)}\delta_s^{k-1}\,ds \int_0^t \mathbb{E}(|\gamma_s^{k-1}|^2)/\delta_s^{k-1}\,ds\\
&\quad\le d^2\int_0^t \lambda_+^k(x_s)\,ds\int_0^t f(s)\,ds \le (d^2+d^3t)e^{(d+1)t}\delta_t^k\int_0^t f(s)\,ds
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left(\left|\int_0^t (\lambda d(x_s^k)^{d-1}+1)\gamma_s^k\,ds\right|^2\right)\\
&\quad\le \int_0^t \left(\lambda d(x_s^k)^{d-1}+1\right)^2\delta_s^k\,ds\int_0^t \mathbb{E}(|\gamma_s^k|^2)/\delta_s^k\,ds\\
&\quad\le (d+1)^2 t e^{(d+1)t}\delta_t^k\int_0^t f(s)\,ds
\end{aligned}
$$

and

$$
\mathbb{E}\left(\left|\int_0^t \gamma_s^{k+1}\,ds\right|^2\right) \le \int_0^t \delta_s^{k+1}\,ds\int_0^t \mathbb{E}(|\gamma_s^{k+1}|^2)/\delta_s^{k+1}\,ds \le e^{(d+1)t}\delta_t^k\int_0^t f(s)\,ds.
$$

Now, from (13), for all $t \le t_0$,

$$
\mathbb{E}(|\gamma_t^k|^2) \le A\delta_t^k + B\delta_t^k\int_0^t f(s)\,ds,
$$

where $A = 4(dt+1)e^{(d+1)t}$ and $B = 8d(d+1)(dt+1)e^{(d+1)t}$. So $f(t) \le Ae^{Bt}$ by Gronwall's lemma. This bound does not depend on ε, so the proposition follows by letting $\varepsilon \to 0$. □

REFERENCES

[1] DARLING, R. W. R. and NORRIS, J. R. (2005). Structure of large random hypergraphs. *Ann. Appl. Probab.* **15** 125–152.

[2] DEIMLING, K. (1977). *Ordinary Differential Equations in Banach Spaces. Lecture Notes in Math.* **596**. Springer, Berlin.

[3] EAGER, D. L., LAZOKWSKA, E. D. and ZAHORJAN, J. (1986). Adaptive load sharing in homogeneous distributed systems. *IEEE Trans. Soft. Eng.* **12** 662–675.

[4] ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York.

[5] GRAHAM, C. (2000). Kinetic limits for large communication networks. In *Modelling in Applied Sciences* (N. Bellomo and M. Pulvirenti, eds.) 317–370. Birkhäuser, Basel.

[6] GRAHAM, C. (2000). Chaoticity on path space for a queueing network with selection of the shortest queue among several. *J. Appl. Probab.* **37** 198–201.

[7] GRAHAM, C. (2004). Functional central limit theorems for a large network in which customers join the shortest of several queues. Preprint.

[8] KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131.

[9] LUCZAK, M. J. (2003). A quantitative law of large numbers via exponential martingales. In *Stochastic Inequalities and Applications* (E. Giné, C. Houdré and D. Nualart, eds.) **56** 93–111. Birkhäuser, Basel.

[10] LUCZAK, M. J. and MCDIARMID, C. (2004). On the maximum queue length in the supermarket model. Preprint.

[11] LUCZAK, M. J. and MCDIARMID, C. (2005). On the power of two choices: Balls and bins in continuous time. *Ann. Appl. Probab.* **15** 1733–1764.

[12] MARTIN, J. B. and SUHOV, Y. M. (1998). Fast Jackson networks. *Ann. Appl. Probab.* **9** 854–870.

[13] MASSART, P. (2002). Tusnády's lemma, 24 years later. *Ann. Inst. H. Poincaré Probab. Statist.* **38** 991–1007.

[14] MITZENMACHER, M. (1996). The power of two choices in randomized load balancing. Ph.D. thesis, Berkeley.

[15] REVUZ, D. and YOR, M. (1991). *Continuous Martingales and Brownian Motion.* Springer, Berlin.

[16] VVEDENSKAYA, N. D., DOBRUSHIN, R. L. and KARPELEVICH, F. I. (1996). Queueing system with selection of the shortest of two queues: An asymptotic approach. *Probl. Inf. Transm.* **32** 15–27.

DEPARTMENT OF MATHEMATICS
LONDON SCHOOL OF ECONOMICS
HOUGHTON STREET
LONDON WC2A 2AE
UNITED KINGDOM
E-MAIL: malwina@planck.lse.ac.uk

STATISTICAL LABORATORY
CENTRE FOR MATHEMATICAL SCIENCES
WILBERFORCE ROAD
CAMBRIDGE CB3 0WB
UNITED KINGDOM
E-MAIL: J.R.Norris@statslab.cam.ac.uk