Subgaussian Concentration and Rates of Convergence in Directed Polymers

Ken Alexander

Univ. of Southern California

Coauthor:

Nikos Zygouras (Warwick)

May 2012

$(1+d)$-dimensional directed polymer in a random environment (DPRE):

Symmetric simple random walk $(x_n)_{n \geq 1}$ on $\mathbb{Z}^d$, distribution P.
Mean-0 disorder $(\omega_{n,x})_{n \in \mathbb{N}, x \in \mathbb{Z}^d}$, i.i.d. in space-time.
Energy associated to a path is the sum of the disorder values along the path. Gibbs measure:

$$dP_{N,\omega} = \frac{1}{Z_{N,\omega}} e^{\beta \sum_{n=1}^N \omega(n,x_n)} dP,$$

where N is the path length and $Z_{N,\omega}$ is the partition function:

$$Z_{N,\omega} = E\left[\exp\left(\beta \sum_{n=1}^N \omega(n,x_n)\right)\right].$$

Here $E[\cdot]$ is over random walk paths.
0-temperature version ($\beta \to \infty$) is last passage percolation.

Fluctuations of $\log Z_{N,\omega}$ and the endpoint x_N:

Exponents χ, ξ defined (loosely) by

$$\left(E_{N,\omega}[|x_N|^2]\right)^{1/2} \sim N^{\xi}, \quad (\mathrm{Var}\,(\log Z_{N,\omega}))^{1/2} \sim N^{\chi}.$$

Gaussian behavior ($\xi = 1/2, \chi = 0$) is known to occur for $d \geq 3$ with small β (Imbrie-Spencer 1988, Carmona-Hu, 2002), but otherwise it is not expected. But we expect subgaussian fluctuations: $\chi < 1/2$. Exponents are believed related by

$$\chi = 2\xi - 1, \tag{1}$$

with values $\chi = 1/3, \xi = 2/3$ for $d = 1$.

Exponent relation (1) confirmed in part, under additional assumptions, through work of Piza (1997) for DPRE, Newman-Piza (1995) and Chatterjee (2011) for first passage percolation.

$d = 1$ values confirmed by Johansson (2000) for 0-temperature (last passage), and by Seppäläinen (2009) for a special disorder distribution.

Fluctuations of $\log Z_{N,\omega}$ and its analogs:

An *exponential bound on scale a_N* means $\exists C_1, C_2$:

$$P\left(|\log Z_{N,\omega} - \mathbb{E}\left(\log Z_{N,\omega}\right)| \geq t a_N\right) \leq C_1 e^{-C_2 t} \quad \text{for all } t > 0, N \geq 1.$$

Kesten (1993): exponential bound on scale $N^{1/2}$ for first passage percolation.
Piza (1997): exponential bound on scale $N^{1/2}$ for DPRE.
Benjamini-Kalai-Schramm (2003): standard deviation $O((N/\log N)^{1/2})$ for first passage percolation.
Benaïm-Rossignol (2008): exponential bound on scale $(N/\log N)^{1/2}$ for first passage percolation.
Seppäläinen (2009): standard deviation of order $N^{1/3}$ for DPRE with $d = 1$ and special disorder distribution.

Free energy of the DPRE:

Straightforward to see that $\mathbb{E} \log Z_{N,\omega}$ is superadditiive in N, so the *free energy* (or *pressure*) exists:

$$p(\beta) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E} \log Z_{N,\omega} = \sup_N \frac{1}{N} \mathbb{E} \log Z_{N,\omega},$$

meaning the discrepancy is nonnegative:

$$s(N) = Np(\beta) - \mathbb{E} \log Z_{N,\omega} \geq 0.$$

The limit is actually an almost sure one (Carmona-Hu 2002):

$$\lim_{N\to\infty} \frac{1}{N} \log Z_{N,\omega} = p(\beta) \quad a.s.$$

Point-to-point version

$$Z_{N,\omega}(x) = E\left[\exp\left(\beta \sum_{n=1}^{N} \omega(n, x_n)\right) \delta_{\{x_N = x\}}\right]$$

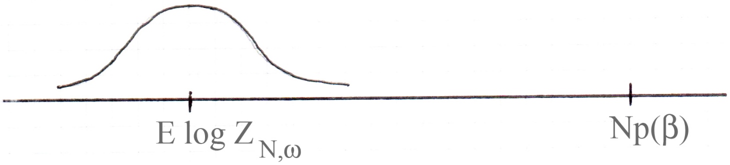
has discrepancy

$$s(N, x) = Np(\beta) - \mathbb{E} \log Z_{N,\omega}(x) \geq 0.$$

At times one needs to understand fluctuations of $\log Z_{N,\omega}$ about its asymptotic approximation $Np(\beta)$, rather than about its mean. Discrepancy has two components:

$$Np(\beta) - \log Z_{N,\omega} = [Np(\beta) - \mathbb{E}\log Z_{N,\omega}] + [\mathbb{E}\log Z_{N,\omega} - \log Z_{N,\omega}]\,. \quad (2)$$

We would like to establish *subgaussian*, or $o(N^{1/2})$, behavior for both of these terms, in the appropriate senses. Methods of A. (1997) show roughly that if the second discrepancy in (2) satisfies an exponential bound on some order a_N, then the first discrepancy is $O(a_N \log N)$. If, as with FPP, we prove the exponential bound on scale $a_N = (N/\log N)^{1/2}$, then this method is not good enough for a subgaussian bound on the first discrepancy.



Log Sobolev inequalities and nearly gamma r.v.'s:

A function f of a gaussian r.v. g (cdf Φ) satisfies a log Sobolev inequality of form

$$\mathrm{Ent}(f) \leq E\left[\left(\frac{df}{dg}\right)^2\right], \tag{3}$$

where $\mathrm{Ent}(f) = \mathbb{E}(f \log f) - (\mathbb{E}f) \log \mathbb{E}f$. For a non-gaussian ω with cdf F, we can express $\omega = T(g)$ with $T = F^{-1} \circ \Phi$ nondecreasing. If for some A we have

$$T'(g) \leq (B + A|T(g)|)^{1/2} \quad a.s., \tag{4}$$

then for functions $f(\omega) = (f \circ T)(g)$, (3) becomes another log Sobolev inequality:

$$\mathrm{Ent}(f) \leq E\left[(B + A|\omega|)\left(\frac{df}{d\omega}\right)^2\right]$$

When (4) (and other minor conditions) hold, we say ω is *nearly gamma*.

Main Results:

Exponential bound on scale $(N/\log N)^{1/2}$:

Theorem 1

Suppose the disorder distribution is nearly gamma and satisfies $E(e^{4\beta|\omega|}) < \infty$. *There exist* C_1, C_2 *such that*

$$\mathbb{P}\left(|\log Z_{N,\omega} - \mathbb{E}\log Z_{N,\omega}| > t\sqrt{\frac{N}{\log N}}\right) \leq C_1 e^{-C_2 t},$$

for all $N \geq 2, t > 0$.

Subgaussian rate of convergence of the mean:

Theorem 2

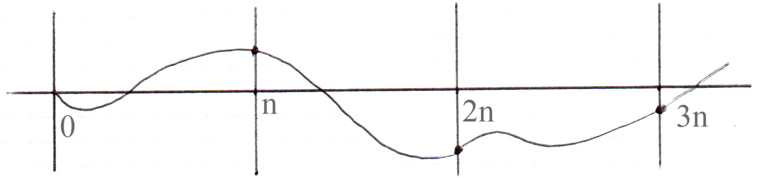
Suppose the disorder distribution is nearly gamma and satisfies $E(e^{4\beta|\omega|}) < \infty$. There exist C_0 such that

$$p(\beta)N - \mathbb{E}\log Z_{N,\omega} \leq C_0 \frac{N^{1/2}\log\log N}{(\log N)^{1/2}}$$

Note the rate is $o(N^{1/2})$ because in contrast to the heuristic from A. (1997), there is a factor of $\log\log N$ in place of $\log N$.

Skeletons: Divide path $(i, x_i)_{i \leq N}$ into blocks of length n, take total length $N = kn$. Consider n fixed, $k \to \infty$.

Points (jn, x_{jn}) form the *simple skeleton* of the path.



There are $O(n^{dk})$ possible skeletons. If there is an event associated to each skeleton which has probability at most $e^{-r_n k}$, then the probability of the union is $O(n^{dk} e^{-r_n k})$, which is small only if r_n is a big enough multiple of $\log n$. This principle is essentially the source of the $\log n$ factor in the A. (1997) hueristic: $\log n$ factor kills off the entropy n^{dk}.

To reduce the required size of r_n to $\log \log n$, we must greatly reduce the entropy to something like $(\log n)^{dk}$, by coarse-graining the skeleton.

Proof elements, exponential bound:

Many ideas from analogous FPP proof of Benaïm and Rossignol (2008).

Let f be a function of the disorder $\{\omega_{i,x} : i \geq 1, x \in \mathbb{Z}^d\}$.

Define $\Delta_{i,x} f = f - \mathbb{E}(f \mid \hat{\omega}_{i,x})$, where $\hat{\omega}_{i,x} = \{\omega_{t,y} : (t,y) \neq (i,x)\}$. This is the average decrease in f when we replace $\omega_{i,x}$ with an independent copy $\tilde{\omega}_{i,x}$ of itself. For f defined in terms of $\log Z_{n,\omega}$, this is related to the Gibbs probability that the path passes through (i,x).

Benaïm and Rossignol (2008): From log Sobolev inequality we get the modified Poincaré inequality

$$\mathrm{Var}(f) \log \frac{\mathrm{Var}(f)}{\sum_{i,x} (\mathbb{E}|\Delta_{i,x} f|)^2} \leq \sum_{i,x} \mathbb{E}\left[(B + A|\omega_{i,x}|) \left(\frac{\partial f}{\partial \omega_{i,x}} \right)^2 \right].$$

To obtain an exponential bound for some function F of the disorder (such as $\log Z_{n,\omega}$), we apply this to $f = e^{\theta F/2}$ with θ small. Sum is over $i \leq n, |x|_1 \leq n$ in that case.

Let
$$\begin{aligned} W_{i,x}(\omega) &= \int \left|f(\omega) - f(\hat{\omega}_{i,x}, \tilde{\omega}_{i,x})\right| \; d\tilde{\omega}_{i,x} \\ &\geq \left| \int (f(\omega) - f(\hat{\omega}_{i,x}, \tilde{\omega}_{i,x})) \; d\tilde{\omega}_{i,x} \right| \\ &= |\Delta_{i,x} f|, \end{aligned}$$

$$W_n(\omega) = \sum_{i \leq n, |x|_1 \leq n} W_{i,x}(\omega),$$

$$r_n = \max_{i \leq n, |x|_1 \leq n} [\mathbb{E}(W_{i,x}(\omega)^2)]^{1/2}, \quad s_n = [\mathbb{E}(W_n(\omega)^2)]^{1/2},$$

$$\ell(n) = \frac{Kn}{\log \frac{Kn}{r_n s_n \log \frac{Kn}{r_n s_n}}} \qquad (K \text{ some constant}).$$

If $r_n = O(n^{-\alpha})$ and $s_n = O(n)$ for some $\alpha > 0$, then $\ell(n) = O(n/\log n)$.

Benaïm and Rossignol (2008) again:

If the right side of the modified Poincaré inequality satisfies

$$\sum_{i,x} \mathbb{E}\left[(B + A|\omega_{i,x}|)\left(\frac{\partial e^{\theta F/2}}{\partial \omega_{i,x}}\right)^2\right] \leq Kn\theta^2 \mathbb{E}[e^{\theta F}]$$

for all $|\theta| < (2\ell(n)^{1/2})^{-1}$, then we have the exponential bound on scale $\sqrt{\ell(n)}$:

$$\mathbb{P}\left(|F - \mathbb{E}(F)| > t\sqrt{\ell(n)}\right) \leq 8e^{-t} \quad \text{for all } t > 0.$$

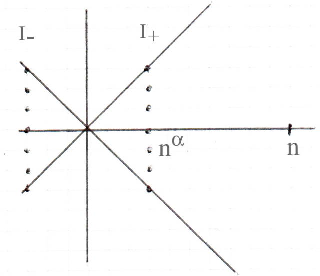
To do this with $\ell(n) = O(n/\log n)$, as noted previously we need good control of the maximum "site influence", in that

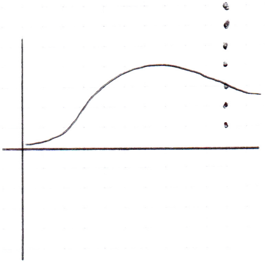
$$r_n := \max_{i \leq n, |x|_1 \leq n} [\mathbb{E}(W_{i,x}(\omega)^2)]^{1/2} = O(n^{-\alpha}) \quad \text{for some } \alpha > 0.$$

We do not have such control for $F(\omega) = \log Z_{N,\omega}$. What to do?

Idea of Benjamini-Kalai-Schramm (2003): Instead of $F(\omega) = \log Z_{N,\omega}$, take $F(\omega)$ to be the average of $\log Z_{N,\omega}$ over some set of translates of the disorder configuration ω—or equivalently, average over different starting points for the path, instead of always (0,0).

Specifically, for $I \subset \mathbb{Z}^{1+d}$ let $\overline{F}^I(\omega)$ be the average of $\log Z_{N,\omega}^{(i,x)}$ over starting sites $(i,x) \in I$; we consider $I = I_\pm$ as shown:





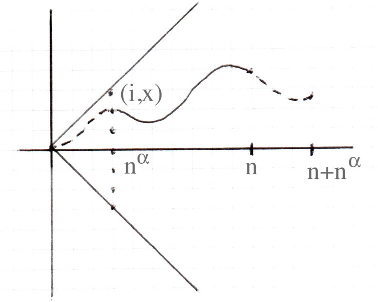
For $\overline{F}^{I_\pm}$, the maximum site influence can be controlled using the fact that for any translate of $I_\pm$, a path can pass through at most one of the sites in the translate.

Therefore we get an exponential bound on scale $(n/\log n)^{1/2}$ for the average $\overline{F}^{I_\pm}$. To turn this into an exponential bound for $\log Z_{n,\omega}$, use bounds of form

$$\log Z_{n,\omega} \geq \log Z_{n,\omega}^{(i,x)} - (\text{error}) \quad \text{for all } (i,x) \in I_+,$$

so, averaging over $(i,x) \in I_+$,

$$\log Z_{n,\omega} \geq \overline{F}^{I_+} - (\text{error}), \quad \text{and similarly} \quad \log Z_{n,\omega} \leq \overline{F}^{I_-} + (\text{error}).$$

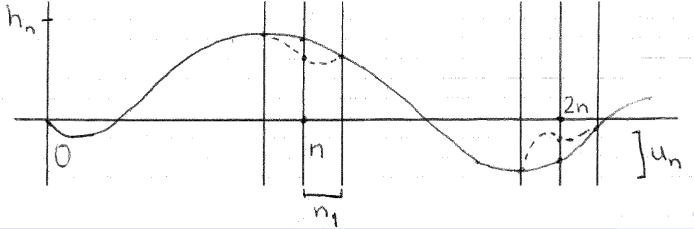


Second proof: size of the discrepancy $Np(\beta) - \mathbb{E} \log Z_{N,\omega}$.

Core idea (which doesn't quite work): *Coarse grained (or CG) skeleton* is skeleton in which all sites $(jn, x_{jn}), j \leq k$, have $x_{jn} \in u_n\mathbb{Z}^d$ for all j where u_n is to be specified; also, in an *acceptable* CG skeleton, transverse increment size $|x_{jn} - x_{(j-1)n}|$ must not exceed h_n (to be specified.) Choose

$$u_n = h_n/(\log n)^3;$$

then there are only $(\log n)^{3d}$ possible increments of a CG skeleton over one block, hence at most $(\log n)^{3dk}$ possible CG skeletons, vs. n^{dk} for simple skeletons. We want to reroute each path, within $n_1 = O(n/(\log n)^3)$ of each hyperplane, so that it passes through at a CG point, and show this does not change $\log Z_{N,\omega}$ too much.



Point-to-point discrepancy

$$s(m,x) = mp(\beta) - \mathbb{E}\log Z_{m,\omega}(x)$$

represents the cost of the increment (m,x). Desired cost level is $O(m^{1/2}\rho(m))$, where

$$\rho(m) = \frac{\log\log m}{C(\log m)^{1/2}}$$

is the factor by which the claimed rate of convergence is subgaussian, that is, we want to show

$$Np(\beta) - \mathbb{E}\log Z_{N,\omega} = O(N^{1/2}\rho(N)). \tag{5}$$

Two cost levels of note:

(m,x) is *efficient* if $s(m,x) \le 4n^{1/2}\rho(n)$ $\qquad (n = \text{block length})$

(n,x) is *adequate* if $s(n,x) \le n^{1/2}(\log n)^{5/2}$.

Note efficient $\leftrightarrow$ subgaussian, but "adequate" does not. Also, if even one (n,x) is efficient, with $n =$ block length, then (5) holds for than n.

Back to core idea: Suppose (to get a contradiction) that no x is efficient. Then for every simple skeleton $\mathcal{S}$ (or CG skeleton $\hat{\mathcal{S}}$) we have

$$\mathbb{E} \log Z_{kn,\omega}(\mathcal{S}) \leq knp(\beta) - 4kn^{1/2}\rho(n)$$

Use exponential bound and the small number of CG skeletons to show that therefore, with high probability,

$$\log Z_{kn,\omega}(\hat{\mathcal{S}}) \leq knp(\beta) - 2kn^{1/2}\rho(n) \quad \text{for every CG skeleton } \hat{\mathcal{S}}; \qquad (6)$$

Use exponential bound again to show that with high probability, for every simple skeleton $\mathcal{S}$ and corresponding rerouted CG skeleton $\hat{\mathcal{S}}$,

$$\log Z_{kn,\omega}(\mathcal{S}) - \log Z_{kn,\omega}(\hat{\mathcal{S}}) \leq kn^{1/2}\rho(n).$$

Unlike (6), here the number of skeletons is not small, but this is compensated by the fact that the rerouted path only differs in a segment of length $O(n/(\log n)^3)$ at each end of each block. Therefore

$$\log Z_{kn,\omega}(\mathcal{S}) \leq knp(\beta) - kn^{1/2}\rho(n) \quad \text{for every simple skeleton } \mathcal{S}.$$

Summing this over all $(2n)^{dk}$ simple skeletons $\mathcal{S}$ shows that, with high probability,

$$\log Z_{kn,\omega} \leq knp(\beta) - \frac{1}{2} kn^{1/2} \rho(n).$$

But this means that with high probability,

$$\frac{1}{kn} \log Z_{kn,\omega} \leq p(\beta) - \frac{\rho(n)}{2n^{1/2}}.$$

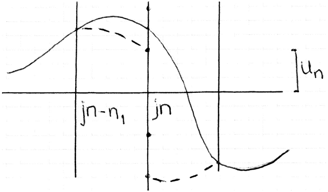
Taking n fixed and k large, this contradicts the a.s. convergence of $\frac{1}{kn} \log Z_{kn,\omega}$ to $p(\beta)$ as $k \to \infty$ with n fixed.

Why doesn't this actually work?

(1) In order to get

$$\log Z_{kn,\omega}(\mathcal{S}) - \log Z_{kn,\omega}(\hat{\mathcal{S}}) \le kn^{1/2}\rho(n)$$

for every simple skeleton $\mathcal{S}$ and corresponding rerouted CG skeleton $\hat{\mathcal{S}}$, we need to know that the rerouted path is not required to make an inefficient increment (one with $s(n_1, x) \gg n^{1/2}\rho(n)$.) Essentially this means the CG spacing u_n must be small enough so that $|x| \le u_n$ implies (n_1, x) is efficient. How can we find such u_n?



(2) We can't require rerouted paths to connect in the hyperplanes. Then after rerouting, CG points in the same hyperplane may be far apart, making the number of possible CG skeletons too large.

(3) We limited the transverse displacement over each block to size $h_n = u_n(\log n)^3$. What about paths/skeletons where larger values occur? Need h_n large here, but that conflicts with need for u_n small.

Idea: Let the system choose h_n, by defining

$$h_n = \max\{|x|_\infty : x \text{ is adequate}\}.$$

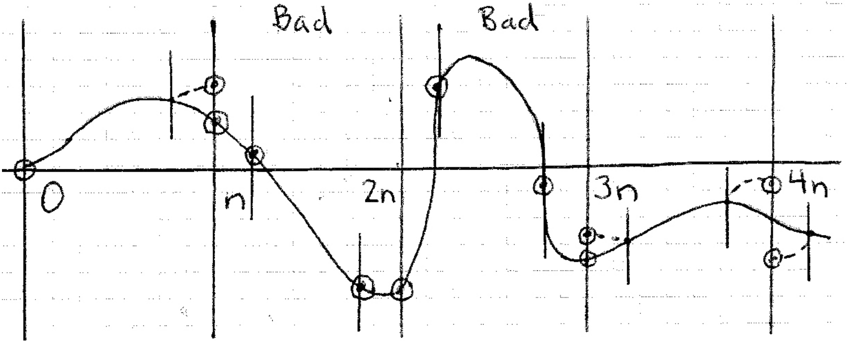
This means any transverse increment greater than h_n is inadequate, i.e. it has a large cost $s(n,x) \geq n^{1/2}(\log n)^{5/2}$. Then take CG spacing $u_n = h_n/(\log n)^3$.

It's not hard to verify that x is adequate whenever $|x|_\infty$ is not too large, so we are not taking the max of the empty set here.

To avoid the other problems, we do not do the CG rerouting in "bad" blocks where either:

(i) the transverse increment over the whole block exceeds h_n, or
(ii) in either of the rerouting intervals $[(j-1)n, (j-1)n + n_1]$ or $[jn - n_1, jn]$, there is a "sidestep," that is, a transverse increment exceeding h_n.

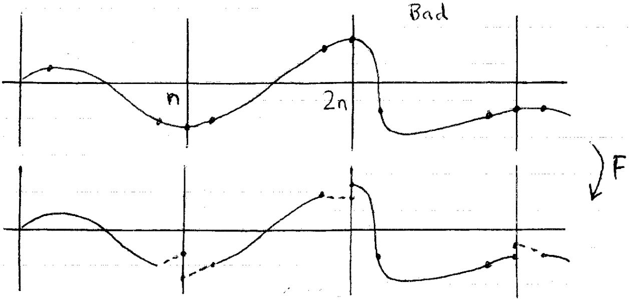
The resulting partial CG approximation:



The *CG-approximate skeleton* $\mathcal{S}_{CG}$ consists of the 2 CG endpoints of the rerouted path in good blocks, and all 4 un-rerouted points in bad blocks.

Large entropy (many skeleton choices) in bad blocks, but this is compensated by the high cost of the inadequate increment, or the sidestep, in a bad block, which can also be shown to have a high cost.

Typical (partially rerouted) “path” corresponding to a CG-approximate skeleton:



To determine the CG-approximate skeleton of a path, it is enough to know the 4 hyperplane intersection points (un-rerouted) in each block (not just bad blocks.) These points form the *augmented skeleton* $\mathcal{S}_{aug}$ of the path, and we have a map $F(\mathcal{S}_{aug}) = \mathcal{S}_{CG}$.

Remaining question: How do we know that when we reroute into a CG skeleton, the rerouted increment doesn't have a much higher cost $s(n_1, x)$ than the original increment? In other words, how do we know that (n_1, x) is efficient for *all* $|x|_\infty \leq u_n$?

Lemma 3

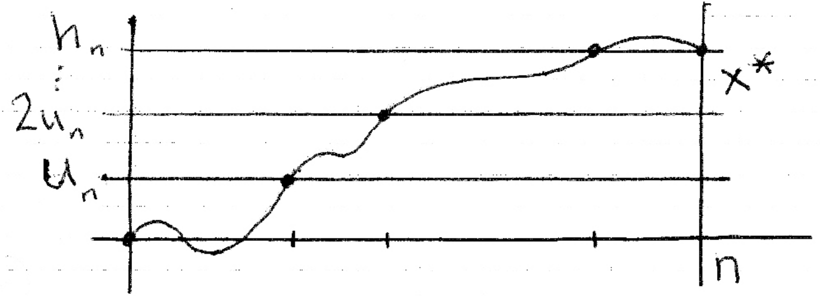
There exists $n_1 \leq 6dn/(\log n)^3$ *such that for all* $x \in \mathbb{Z}^d$ *with* $|x|_\infty \leq u_n$*,*

$$s(n_1, x) = n_1 p(\beta) - \mathbb{E} \log Z_{n_1,\omega}(x) \leq 20dn^{1/2}\rho(n).$$

This means we can always "climb" transversally a distance u_n, to connect to a CG point, at only a small cost.

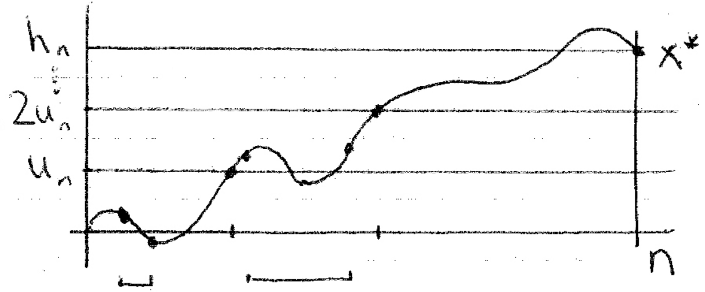
Proof idea for Lemma 3:

Recall that $h_n = \max\{|x|_\infty : x \text{ is adequate}\}$. This means a path can climb a distance $h_n = (\log n)^3 u_n$ at only moderate cost. Let x^* be adequate with $|x^*|_\infty = h_n$, say $x_1^* = h_n$. For a path $\gamma = (i, x_i), i \le n$ from $(0,0)$ to (n, x^*), the *climbing skeleton* $\mathcal{S}_{cl}(\gamma)$ consists of the first points where heights $ju_n, j \le (\log n)^3$, are reached. *Climbing segments* are the segments between such points.



Climbing segments are *long* or *short* according as their length exceeds $2n/(\log n)^3$. (Twice the average length.) At least half are short.

A climbing segment is *clean* if the increment of every sub-segment is efficient. A path is *totally soiled* if no short climbing segment is clean. A totally soiled path (underlined segments are inefficient):



Lemma 4

Provided n is large, there exists a path γ from $(0,0)$ to (n, x^*) containing a short climbing segment which is clean. (That is, γ is not totally soiled.)

Lemma 5

Provided n is large, there exists a path γ from $(0,0)$ to (n, x^*) containing a short climbing segment which is clean. (That is, γ is not totally soiled.)

Let

$$\mathcal{D}^* = \{\text{all totally soiled paths from } (0,0) \text{ to } (n, x^*)\}.$$

The lemma follows if we show

$$\mathbb{P}(Z_{n,\omega}(\mathcal{D}^*) < Z_{n,\omega}(x^*)) > 0. \qquad (7)$$

Since x^* is adequate, we have

$$\mathbb{P}\left(\log Z_{n,\omega}(x^*) > p(\beta)n - 2n^{1/2}(\log n)^{5/2} \right) > \frac{1}{2}.$$

Totally soiled paths have many inefficient increments, so we can show

$$\mathbb{P}\left(\log Z_{n,\omega}(\mathcal{D}^*) > p(\beta)n - 2n^{1/2}(\log n)^{5/2} \right) \ll \frac{1}{2}.$$

Together these prove (7).

Our short clean climbing segment α climbs a distance u_n in a time at most $2n/(\log n)^3$, and every subincrement is efficient. We can show that with these properties, for any x with $|x|_\infty \leq u_n$, one can cut and paste at most $4d+1$ (necessarily efficient) segments of α to assemble a path from $(0,0)$ to (n_1, x):

$$(n_1, x) = \sum_{i=1}^{4d+1} \big[(m_i, z_i) - (l_i, y_i)\big],$$

where α passes through the points (m_i, z_i) and (l_i, y_i). This is enough to prove Lemma 3, since the cost $s(\cdot,\cdot)$ is subadditive.