Update

Pete Rigas, Lambert Lab

March 8, 2021

dirac delta landscape potential

Such a class of potentials

- is resemblant of the energy barriers of binding,
- composed of step functions with a fixed "jump" in the energy landscape between neighboring base pairs,
- produces solutions $\mathcal{S}$ that are dependent on the magnitude of the jump discontinuity of the landscape between neighboring base pairs, taking the form

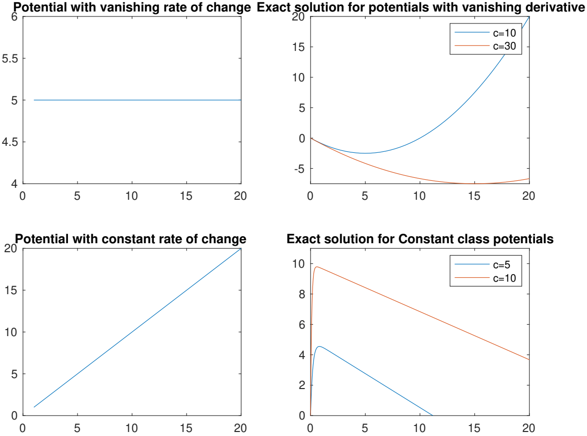
$$\mathcal{S}(x) = \frac{x}{|\mathcal{U}'(x_1)|} - \frac{\exp(|\mathcal{U}'(x_1)|\ x)}{|\mathcal{U}'(x_1)|^2} + \frac{1}{|\mathcal{U}'(x_1)|^2}\ ,$$

  where the landscape $\mathcal{U}'(x)$ is evaluated at the barrier corresponding to base pair x_1 in the sequence,
- is problematic to code in Matlab.

code

```
function [sol_array] = step_potential(U_initial, U_string)
sol_array = [ ];
% solve IVP for first barrier
syms y(x)
Dy = diff(y);
ode = diff(y,x,2) == -1+(U_initial)*diff(y,x,1);
cond1 = y(0) == 0;
cond2 = Dy(0) == 1;
conds = [cond1 cond2];
ySol(x) = dsolve(ode,conds);


for I = 1 : length(U_string)

    temp = matlabFunction(ySol(x));

    % formulate new IVP from exit time of previous solution
    ode_temp = diff(y,x,2) == -1+(U_string(I))*diff(y,x,1);
    cond_1 = temp(1);
    cond_2 = Dy(0)==1;
    conds_temp = [cond_1 cond_2];
    ySol(x) = dsolve(ode_temp,conds_temp);

end


end
```

observing deviations in exit time from other classes of potentials



objective: observe similarly predictable fluctuations in predicted exit times for the dirac delta landscape

data pipeline

We construct, and output,

- solutions for the IVP, exit time approximations, from distributions prepared over each base pair,
- variation in the approximation of exit time, across base pairs of the target sequence,
- relation between barrier height and approximation of exit time,
- approximations of exit time through evaluations of closed solutions.

preparing Gaussian distributions over the landscape

- Step 0: Set the mean of each Normal distribution halfway through the inspection step, at each base pair.
- Step 1, adjust peak of Gaussian pmf over each landscape barrier: $p(x) = \frac{\exp(-(x-\mu)^2)}{\sqrt{2\pi\sigma}} + h = \frac{\exp(-(x-\mu)^2) + \exp(\log\sqrt{2\pi\sigma h^2})}{\sqrt{2\pi\sigma}}$, which can be further rearranged to evaluate with three Matlab function handles, including
  - evaluation of an exponential term dependent on the mean and variance,
  - adding 1 to the result above,
  - multiplication by an evaluation of the standard pmf without the horizontal shift,

  from the finalized Gaussian, $p(x) = \frac{\exp\{-(x-\mu)^2\}}{\sqrt{2\pi\sigma}}\left(1 + \exp\{\log\sqrt{2\pi\sigma h^2} + (x-\mu)^2\}\right)$, and $h = \mathcal{U}'(x_{i+1}) - \mathcal{U}'(x_i)$.

- Step 2, numerically approximate exit time through IVP: Solve the second order dimensionless equation with $\mathcal{U}'(x) = p(x)$, $\tau^i(0) = 0$ if $i = 1$ and $\tau^i(0) = \tau^{i-1}(1)$ for $i \geq 2$.
- Step 3: Quantify deviations in the exit time from constant $\mathcal{U}'$, from the following numerical instances,
  - height h of the energy barrier approaching infinity,
  - translating the mean to the left or right of the middle of the base pair,
  - decreasing the magnitude of the variance of the distribution.
- Step 4 (towards the ultimate goal): Numerically characterize dependence of the rate of change in the exit time τ with the fatness of the tails of the distributions constructed from Steps 0-3.

Remark I: We make use of relation $\max_x\{p(x)\} \approx \frac{0.4}{\sigma}$, eliminating the horizontal shift h to avoid issues with implementing function handles of multiple variables.

components of the solution at each step

Across every base pair of the target sequence, the solution consists of:

- (I) an exponential whose power is dependent on the error function,
- (II) the square of the integral of the same exponential, whose region of integration is determined by the base pair length of protein inspection,
- (III) a contribution from a definite integral over the number of base pairs inspected,
- (IV) a trapezoidal approximation of the exponential, dependent on the variables of integration of (III),

from which the solution is proportional to the contribution $(\mathrm{I}) + (\mathrm{II}) + (\mathrm{III})(\mathrm{IV})$.

organizing each iterate

Our loop indices are,

- the amplitude of the landscape barrier,
- the mean and variance of the prepared distribution,

from which,

- IVP is solved forwards in time, after inspection of one base pair, with the landscape $\mathcal{U}'(x)$ given by the prepared distribution,
- approximations of exit time are obtained.

Remark II: To avoid making use of the vertical shift h, we choose sufficiently small $\sigma' < \sigma$, so that

$$\sigma' \approx \sigma + h \Leftrightarrow \frac{0.4}{\sigma'} \approx \frac{0.4}{\sigma + h} \ .$$

simulation test cases

Solve the second order exit time problem subject to:

- Case I: Gaussian potentials with mean concentrated at the middle of each base pair in inspection, with increasing variance, for the nonconstant derivative coefficient,
- Case II: Gaussian potentials with the same mean, with variance across base pairs of the sequence that is not necessarily montonically increasing,
- Case III: 'Tweaking' the variance of the distributions more deliberately at different base pair resolutions.

```
printing first term
    3.45345021056028e+000

COMPLETE FIRST TERM
    989.933492274505e-003

int(exp(-(2251799813685248*pi^(1/2)*erf(u - 5/2))/3192969000987793), u, 0, x
string
second integral term
    11.9263183568189e+000

FIRST AND SECOND TERMS
   -10.9363848645443e+000

double integral term
1963889140381397/562949953421312

exp(-(2535301200456458599293666380957*erf(u/2 - 5/2))/2028240960365167042394725128601 6)

exp(-(2535301200456458599293666380957*erf(u - 5/2))/20282409603651670423947251286016)

    @(u)(u.*exp(sqrt(pi).*erf(u-5.0./2.0).*(-7.052369794346954e-1)).*(exp(erf(u-5.0./2.0).*(-1.25)).*3.488567906340373-exp(erf(u./2.0-5

    2.85684899075369e+000

FIRST, SECOND AND THIRD TERMS
    5.85684899075369e+000

end
    5.85684899075369e+000
```

Figure 1: execution

Please let me know how to embed the video output for the March Meeting? I tried using the media9 library.

figure, subplots with the exit time plotted against the base pair, and the landscape potential at every base pair of the sequence

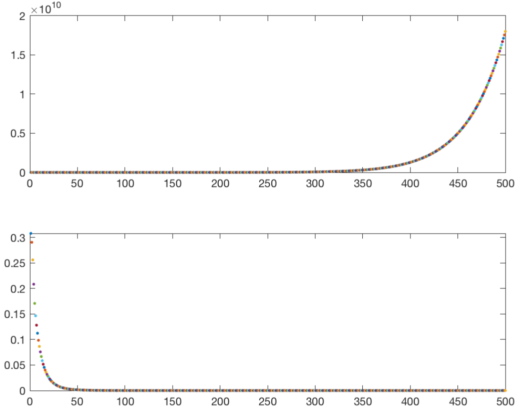


Figure 2: Exit time, *top*, reaction time, *bottom*. 500 sequences sampled, $\max \sigma_1 \approx \frac{0.4}{500} \approx 0.008$.

rerunning the simulation for higher order variances in the Gaussian noise

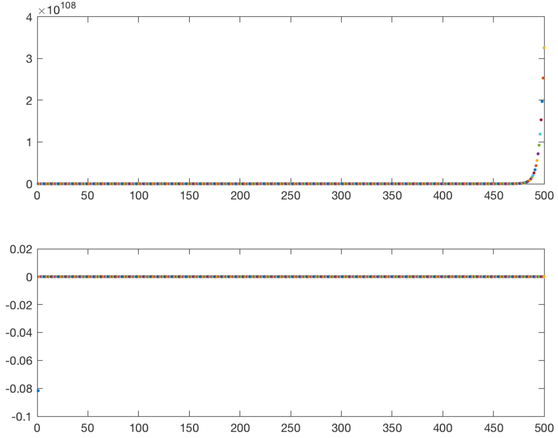


Figure 3: Exit time, *top*, reaction time, *bottom*. 500 sequences sampled, $\max \sigma_2 \approx \frac{0.4}{500 \times 100} \approx 0.000008$.

drawbacks

Despite obtaining exit times of arbitrarily large order as $\sigma \longrightarrow 0^+$,

- the simulation drawn under σ_1 asymptotically approaches the largest exit time at the 500th base pair of the sequence in comparison to the simulation drawn under σ_2 ,
- becomes numerically unstable due to issues of incorporating the initial condition at the starting point of time evolution (dsolve does not take integer inputs for initial conditions),

leading to modifications in which,

- the landscape is partitioned over base pairs in which it is expected that the exit time substantially decrease in magnitude corresponding to dramatic changes in the magnitude of the barrier,
- the distribution of exit times is translationally invariant, and determined by the location of barrier height magnitude in the target sequence.

preparing test cases

We also modify the potential construction and interval over which the solution is obtained with each iterate, making use of

- $\prod_i \frac{1}{\sqrt{2\pi\sigma_i}}$,
- $\prod_i \exp(-(x-\mu_i)^2)$,
- the portion of the target sequence that is inspected, for which we can always use the initial condition $\tau(0) = 0$, to avoid errors with generating initial conditions with each additional time step in dsolve.

In terms of test cases, we maintain the 500 bp target sequence, in which we test

- target sequences whose potential, per base pair inspected, falls into the cases given on the next slide.

variance test cases

Case	σ_i
1	$\sigma_1 = 8000 - (1 \times 10^{-15})x - x^2$
2	$\sigma_2 = 8000 - (1 \times 10^{-15})x - x^2 + (5 \times 10^{-4} x^3) + 0.09x^4 -$
2	$(7 \times 10^{-8})x^5$
3	$\sigma_3 = \sigma_2 + 50x^6$
4	$\sigma_4 = \widetilde{\sigma_2} + 50\log(x^6)x^6$

For more nontrivial landscapes, the simulation is modified across the identical sample space of sequences, as the variance with respect to the base pair varies.

1

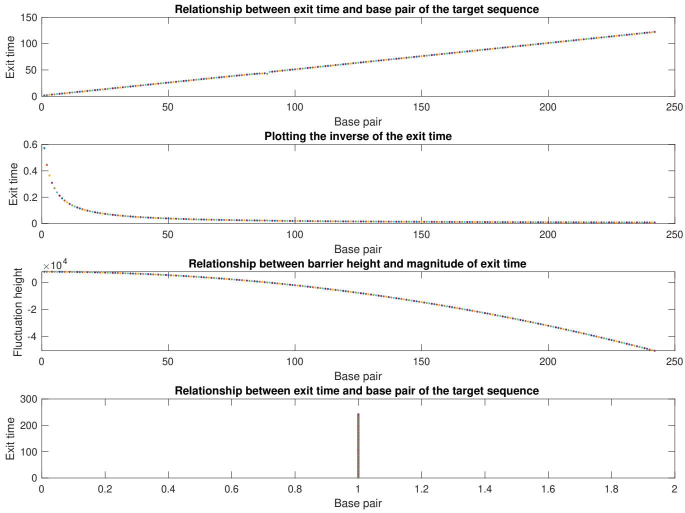


Figure 4: σ_1

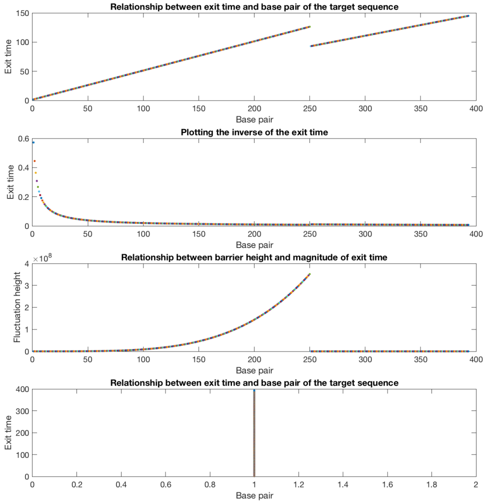


Figure 5: σ_2

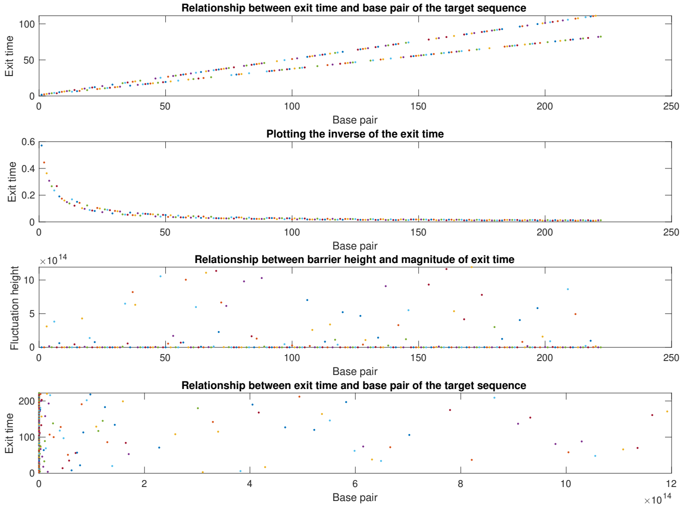


Figure 6: σ_3

4

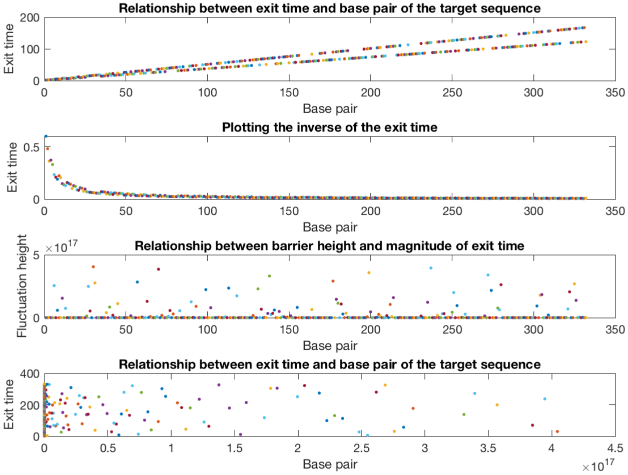


Figure 7: σ_4