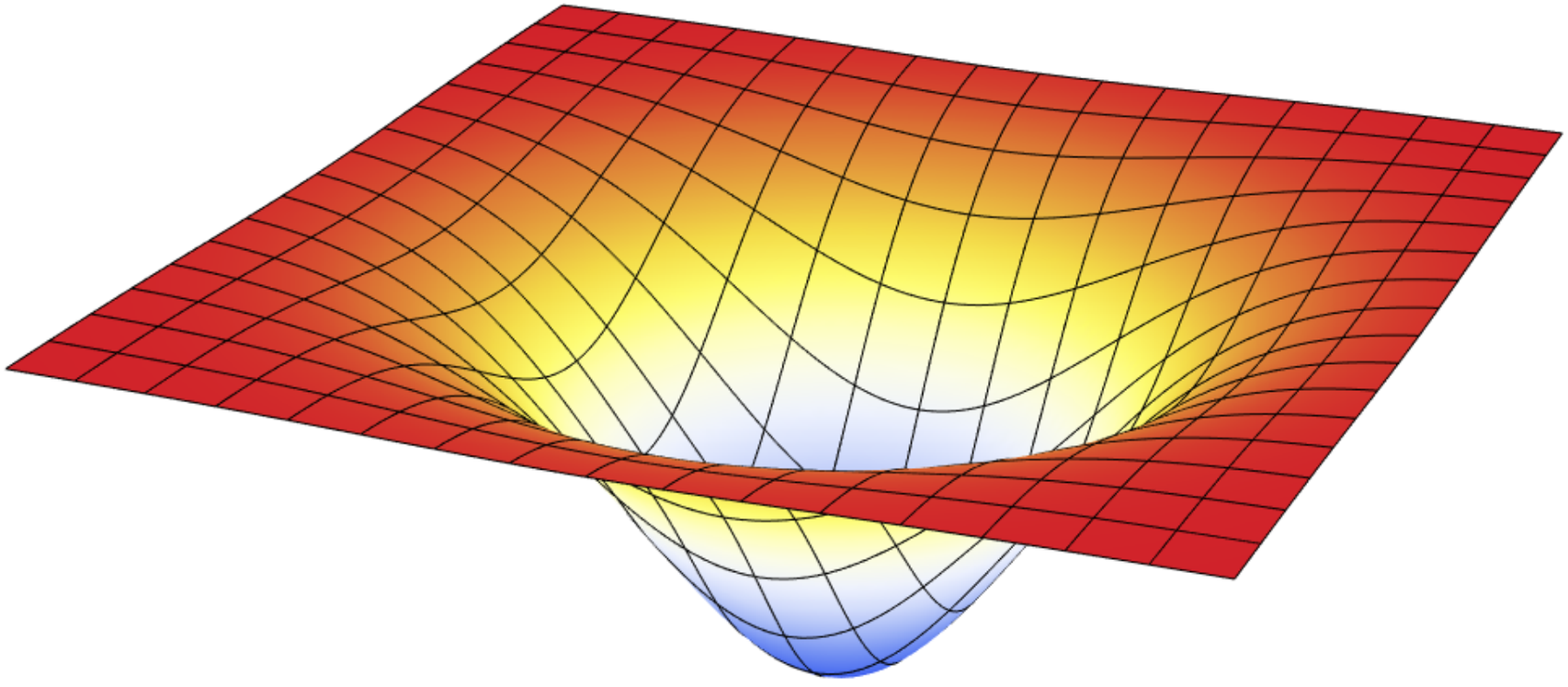
DNA elastic energy minimization

emDNA presentation



Overview

- DNA mechanics
  - geometry
  - elastic energy
  - kinematical constraints
- Minimization crash course
  - methods and limitations
  - application to DNA elastic energy
- *emDNA* software
  - presentation
  - data format and options
  - sequence-dependent elasticity
  - examples
- Tips
  - log file description
  - common mistakes

All materials will be available:
slides, method paper, library interface description, and software

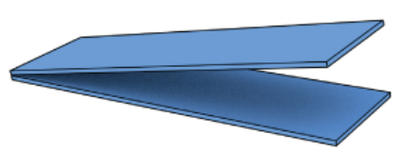
myosin:/home/shared/emDNA

DNA mechanics - geometry

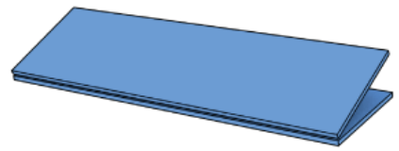
description of a single step

6 rigid-body parameters
=
3 angles
+
3 vector components

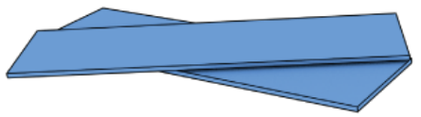
$\underline{p} = (\theta_1, \theta_2, \theta_3, \rho_1, \rho_2, \rho_3)$



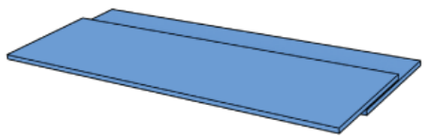
tilt θ_1



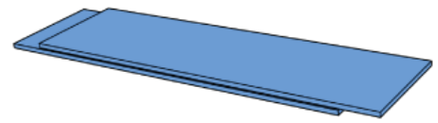
roll θ_2



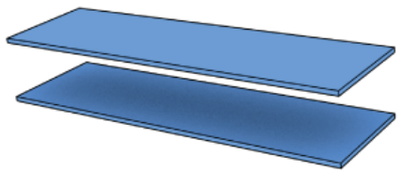
twist θ_3



shift ρ_1



slide ρ_2

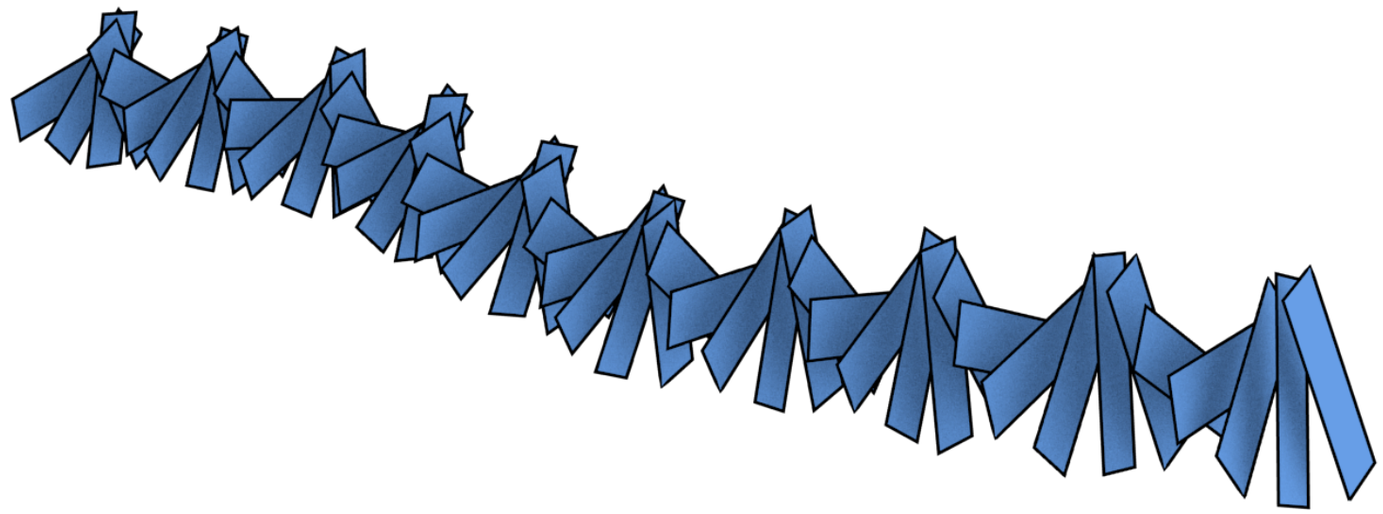


rise ρ_3

description of a collection of base pairs

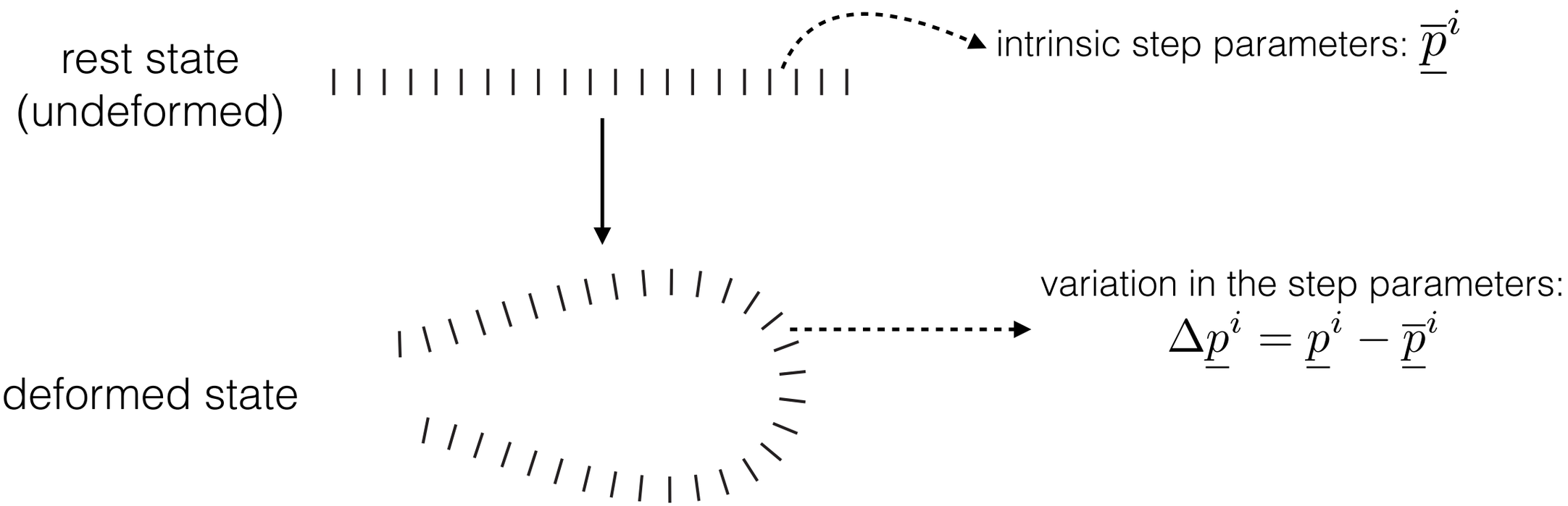
N base pairs
=
6(N-1) parameters

$\underline{P} = \{\underline{p}^i\}$



DNA mechanics - elastic energy

The elastic energy of DNA measures the *work* associated with the deformation (*i.e.*, change in the geometry) of base-pair steps.



total elastic energy
defines a sum over all steps:

$$\mathcal{E} = \frac{1}{2} \sum_{i=1}^{N-1} \Delta \underline{p}^{i\intercal} \boldsymbol{F}^i \Delta \underline{p}^i$$

the sequence-dependent elasticity
is encoded in:

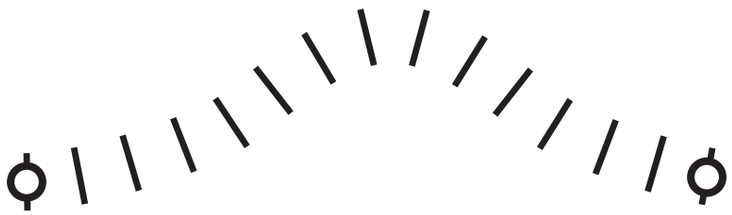
$\underline{\overline{p}}^i$ intrinsic step parameters

$\boldsymbol{F}^i$ force constant matrix (6 x 6)

DNA mechanics - kinematical constraints

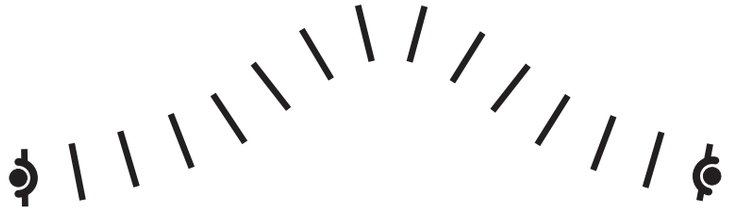
What are the possible constraints ?
The constraints correspond to the end conditions applied to the first and last base pairs.

no end condition
(free collection)



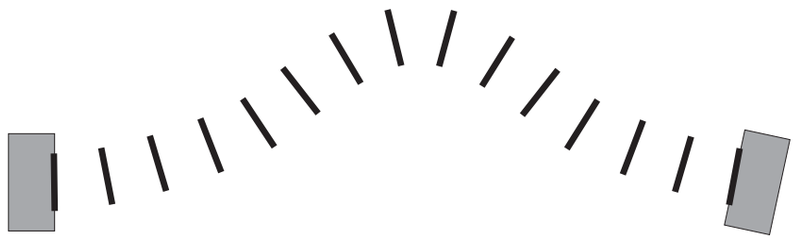
no constraints

first and last
origins imposed



fixed end-to-end vector

first and last
base pairs imposed



fixed end-to-end vector
and
end-to-end rotation

We can also mix these conditions:
for example, the end-to-end vector can be constrained along an axis.

What we want to do …

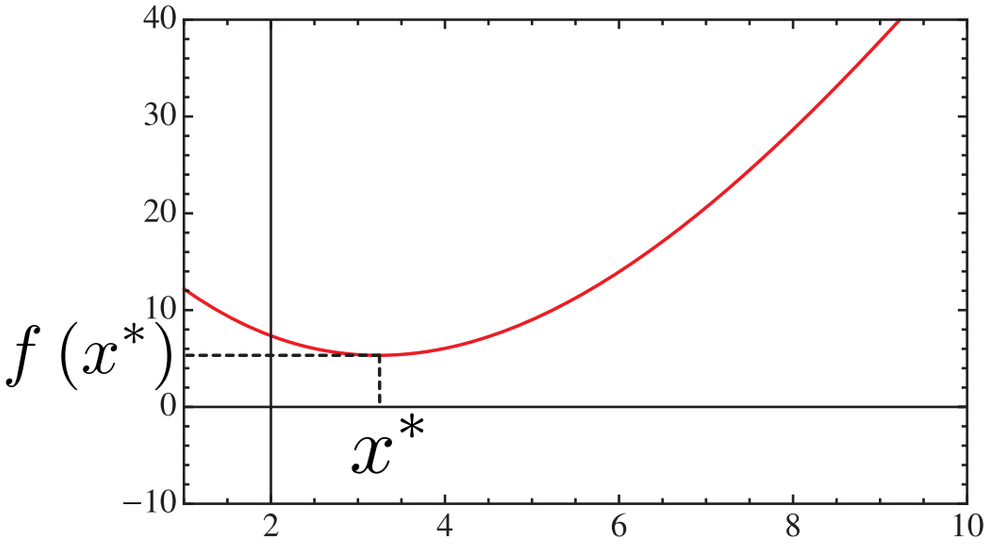
We want to find the optimal configuration of
a collection of base pairs with given end conditions.

[in maths language]

We want to find the set of step parameters
that minimizes the total elastic energy (*i.e.*, the deformation)
under given boundary conditions.

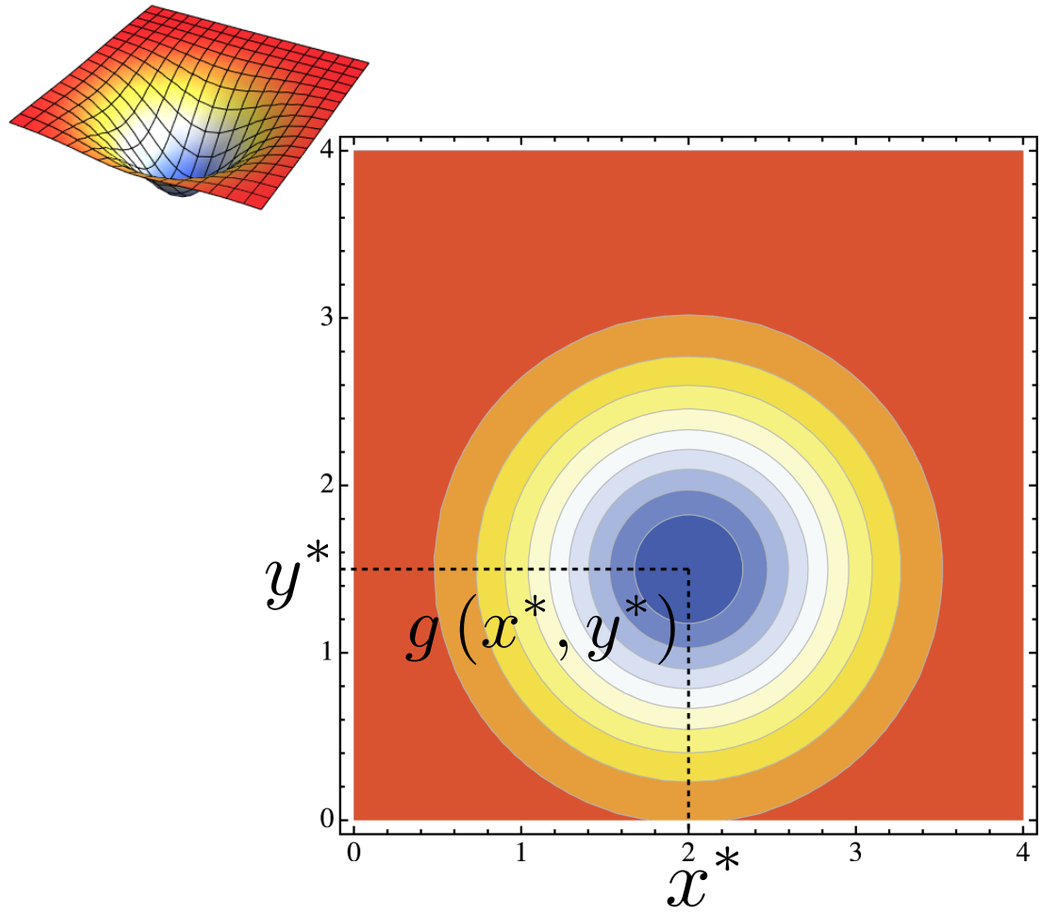
Minimization - introduction

We want to find the minimal value(s) of a given function.



$$\min_{x \in \mathbb{R}} f\left(x\right) = f\left(x^*\right)$$

$$f'(x^*) = 0$$

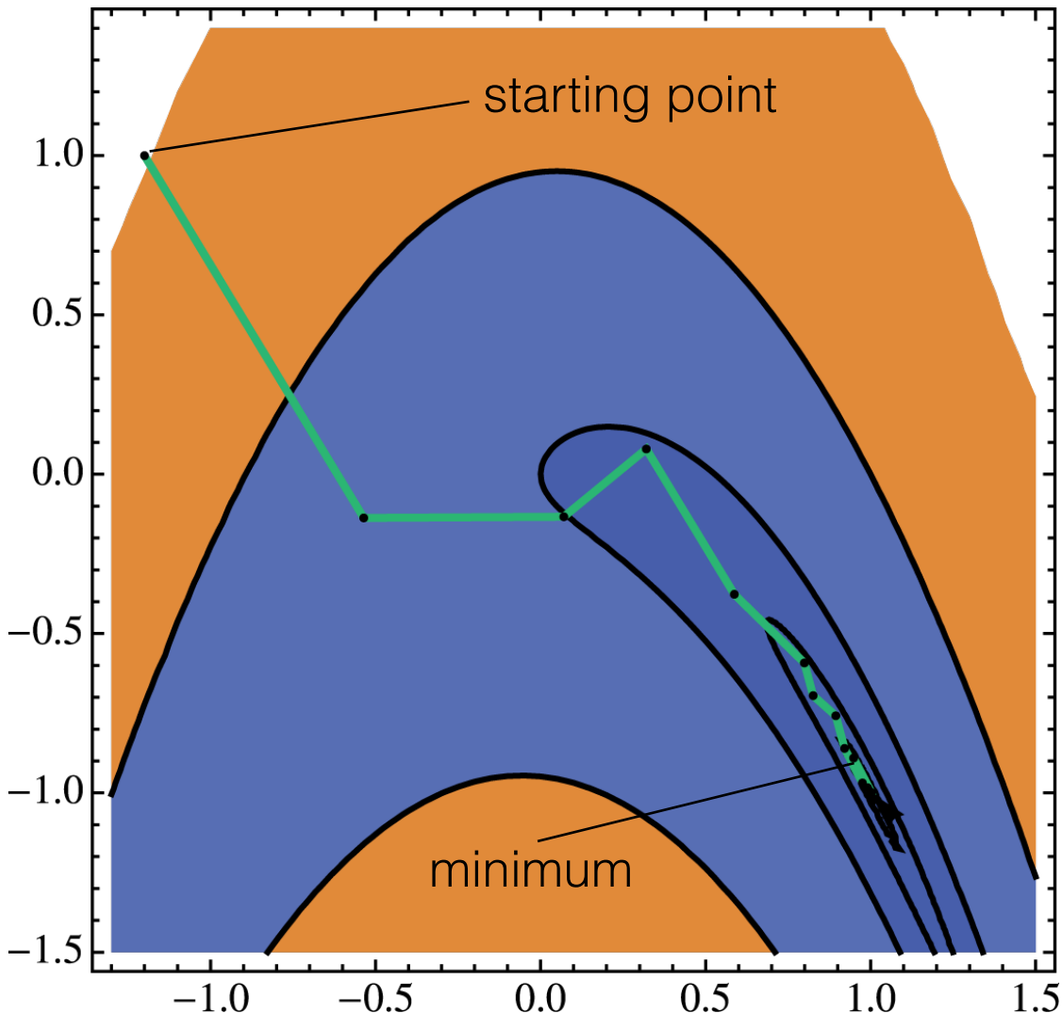


$$\min_{(x,y) \in \mathbb{R}^2} g\left(x, y\right) = g\left(x^*, y^*\right)$$

$$\left.\frac{\partial g}{\partial x}\right|_{x=x^*} = 0 \qquad \left.\frac{\partial g}{\partial y}\right|_{y=y^*} = 0$$

Minimization - methods

Most minimization methods requires the derivatives of the objective function.
The idea is to follow the direction along which the function decreases.



<u>general approach:</u>

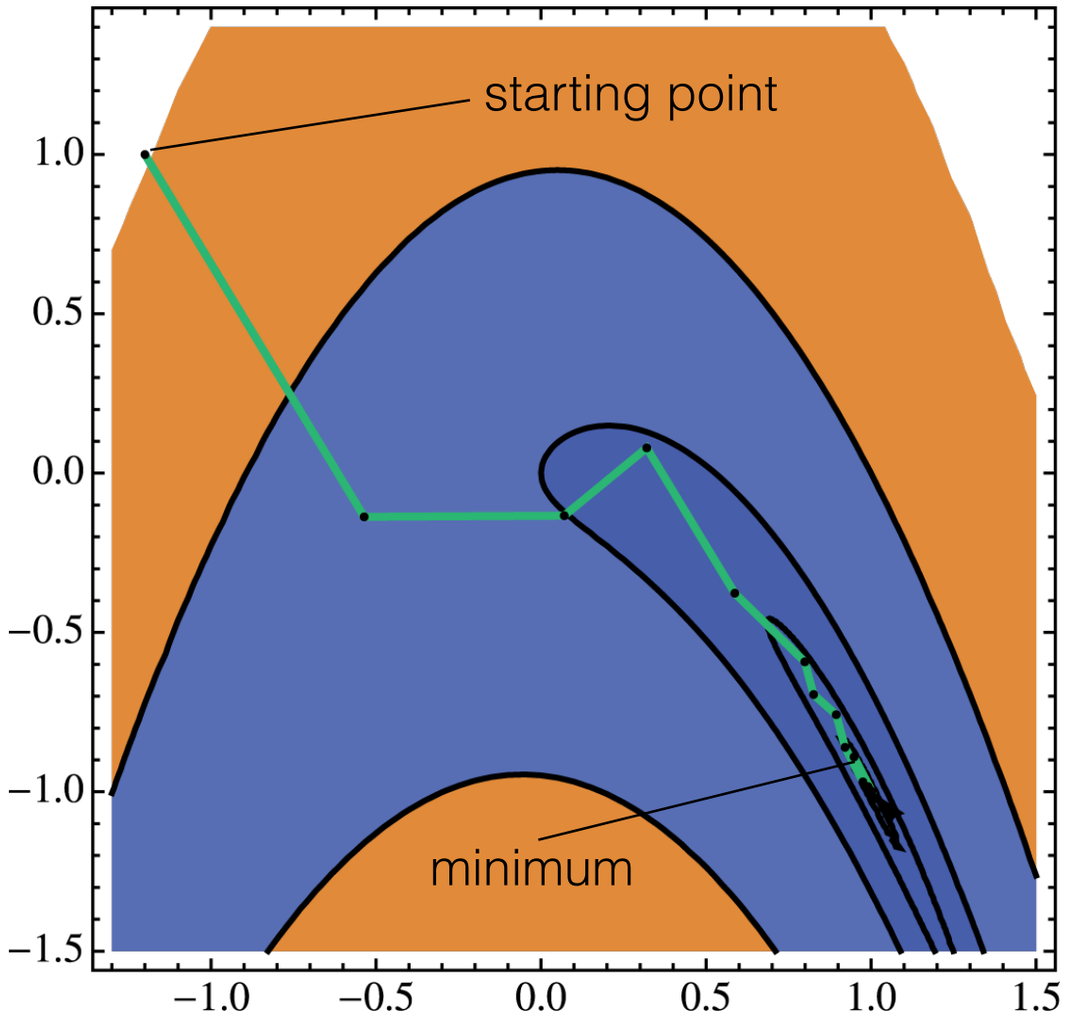
- pick a starting point,
- follow the gradient direction (*descent*),
- stop after some steps (*criteria*),
- compute the gradient at the new point,
- keep going on for many iterations…,
- stop when the gradient norm is zero.

<u>algorithms:</u> gradient descent, conjugate gradient (CG), L-BGFS, … all rely on the same general idea.

It almost looks to good to be true.

Minimization - limitations

There are a lot of limitations for all minimization methods.



<u>starting point:</u>

what is the influence of the starting point ?

<u>descent along the gradient:</u>

how many steps? what about the step size ?

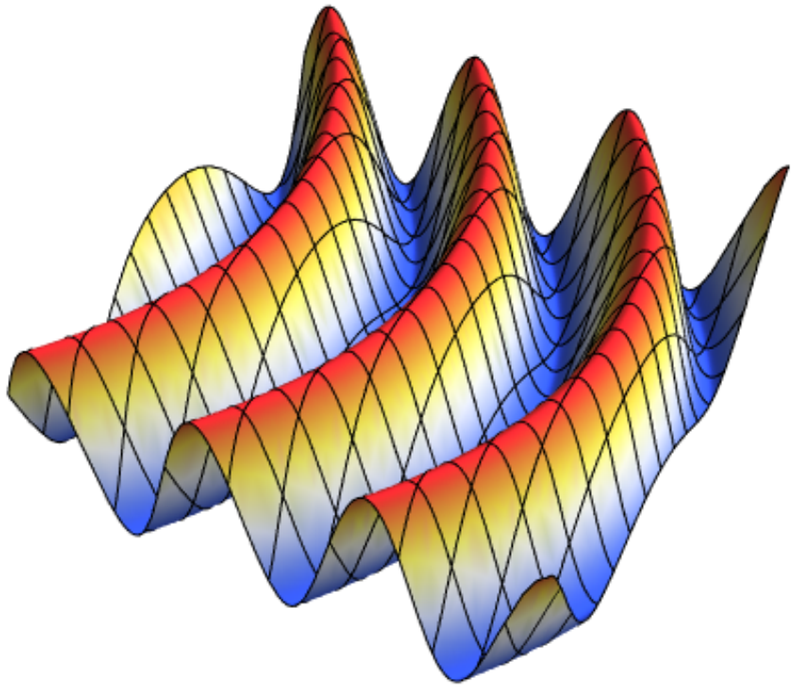
<u>iterations / stopping condition:</u>

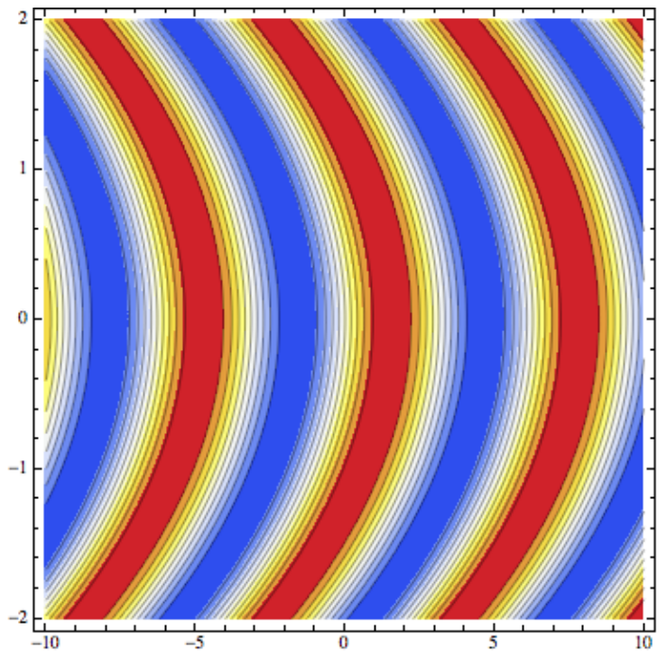
- how many iterations should we perform ?
- on a computer the gradient norm will never be zero … we need a threshold.

But that's not all …

Minimization - limitations

What if we try to minimize a “weird function” ?
weird = irregular, multiple minima, …





<u>starting point:</u>
depending on where we start we will not arrive at the same minimum.

<u>descent along the gradient:</u>
if the step size is too large we might *jump over valleys.*

<u>iterations / stopping condition:</u>
the gradient at the peak of a crest has a norm equals to zero.

And more importantly …
To date, there is no minimization method that guarantees to find a global minimum, so we are stuck with local minima.

Minimization - DNA elastic energy

In spite of all those limitations, it (can) work

…

if we take some precautions.

<u>starting point:</u>

prepare a *good* guess, that is, something reasonable with respect to the boundary conditions.

<u>descent along the gradient:</u>

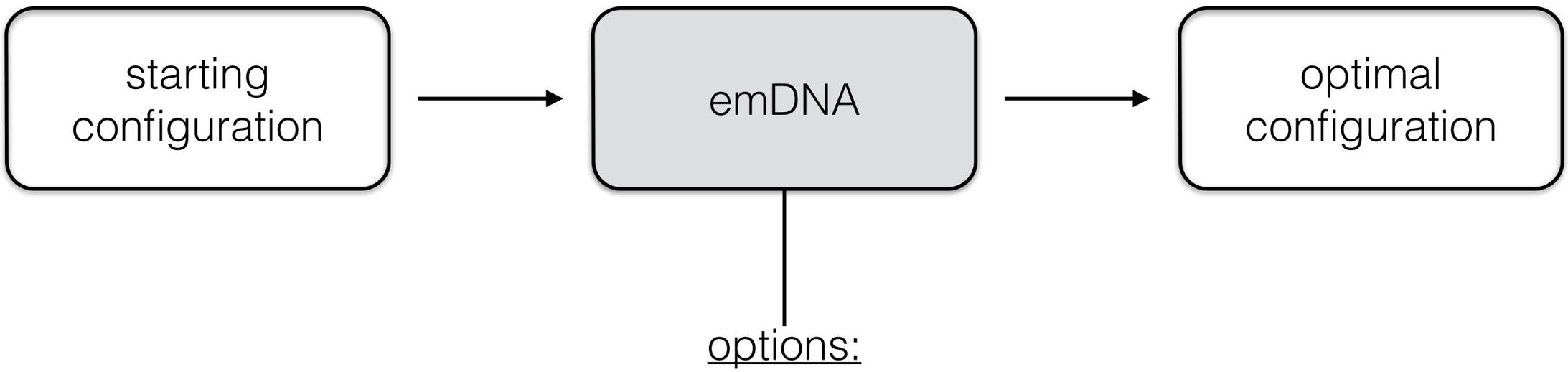
use conservative step size (*i.e.*, small) and adjust it through multiple runs if needed.

<u>iterations / stopping condition:</u>

once an optimal solution is found add a random perturbation to it and minimize again; if it converge to the same solution you are good to go,

emDNA - presentation

emDNA is a software for minimizing the elastic energy of
a collection of base pairs subjected to end conditions.



companion tools:

emDNA_parser

data format converter

emDNA_topology

topological numbers computation

emDNA - data format

There are three different formats to describe the configuration of a collection of base pairs.

bplist

- simple format with no sequence information,
- each line corresponds to a base-pair frame.

example:
{{origin_x, origin_y, origin_z}, {{d1x,d1y,d1z},{d2x,d2y,d2z},{d3x,d3y,d3z}}}

x3DNA - base pairs OR step parameters

- same format as x3DNA,
- contains sequence information.

example:
[[see x3DNA documentation]]

emDNA - sequence-dependent elasticity

There are several models for the force field.
Force field = intrinsic step parameters + force constant matrix.

The command line options is:
`--DNA-seqdep-model=[string]`

IdealDNA:
- no sequence dependence,
- B-DNA intrinsic step parameters with an helical repeat of 10.5 turns,
- isotropic force constants with traditional bending and twisting moduli (quasi-inextensible).

Olson1998:
- full sequence dependence,
- intrinsic step parameters extracted from crystal structures,
- force constants obtained by covariance analysis.

AnisoDNA:
- same as IdealDNA but with anisotropic force constants.

For more details look into the headers file! (/home/shared/emDNA/install/include/emDNA/…)

emDNA - command line options

All the options can be listed with the command (works for the tools too):

`./emDNA --help`

Most common options:

Input / Output:

- input file: `--bp-list-input=…`, `--x3DNA-bp-input=…`, `--x3DNA-bp-step-params-input=…`,
- output files prefix: `--output-name=myminim`,
- progress indicator: `--energy-progress`,
- output configurations during minimization: `--output-progress=every_n_steps`.

End conditions (exclusive options):

- free collection: `--free-collection`,
- imposed end-to-end vector: `--hold-last-origin`,
- imposed end-to-end vector+rotation: `--hold-last-bp`.

Others:

- settings: `--minim-settings="{max_iterations, dx, df, dg, step-size}"`.
- protein constraint (freeze the step parameters over a range):
  `--frozen-steps="{n_1:n_2,m_1:m_2}"`.

Avoid the other options!

emDNA - command line output

Example:

```
[iMac] Release > ./emDNA --bp-list-input=minicircle_104bp.txt --hold-last-bp --output-name=minicircle_104bp --DNA-seqdep-model=IdealDNA
--- bp collection input
  bp collection created from input file: minicircle_104bp.txt
  bp colllection size: 105-bp (104 steps)
  DNA sequence-dependence model: IdealDNA
  bp collection frozen steps:
--- minimization setup
  minimization type: Alglib gradient-based minimization - last bp fixed

--- minimization results
  energy initial value: 26.9428078914
  energy final value: 26.9417651311
  gradient norm angle: 0.0040576194
  gradient norm distance: 0.0000646159
  # iterations: 62
  return code: EPSG

[iMac] Release >
```

checklist:

- energy initial and final values,
- gradient norms (mind the rescaling!),
- return code!

return code:

- EPSG: gradient norm is less than 10-4,
- EPSF: function is not decreasing enough (step size!),
- EPSX: point is not changing enough (step size!),
- MAXIT: max # iterations reached,
- anything else … you are in trouble.

emDNA - output data

emDNA creates several output files when the minimization is complete.
The files are named `emDNA_minim*`
(this can be changed with `-output-name=MyPrefix`).

emDNA_minim.log	Log file containing details about the minimization (to be checked!)
emDNA_minim_opt.txt	File containing the optimal collection of base pairs in the same format as the input.
tmp_confs.txt	File containing the configurations if `--output-progress` is passed on the command line. Format: a line is a list of frames (similar to bplist).

emDNA - play time!

Tips - log file description

The log file contains a bunch of information about the minimization run.

values of all the command line options

detailed information about the starting collection

detailed results about the minimization

energy repartition among the different modes of deformation

Tips - common mistakes

<u>clean input:</u>

whatever format you use, make sure that it does not contains any scientific notation (plain numbers only):

2E-2, 3*10^-9, …

<u>orthogonalized base pair frames:</u>

orthogonalize your base-pair frames in the input configuration

…

otherwise you might have some surprises.

<u>precision does come out of nothing:</u>

if your input has a precision of N digits do not expect to get result with a better precision (or at least, do not believe anything after N digits).