MATH590 DIF: Information geometry

In this final module homework we combine concepts of differential geometry with information theory to investigate the utility of the new field of information geometry to data analysis. We'll use the same model system of viscoelastic flow in a convergent channel. Recall that this simple physical system corresponds to a generic situation encountered across the sciences: a system subjected to background stochastic forcing and experiences a change in external forcing (i.e., the channel wall geometry in this case).

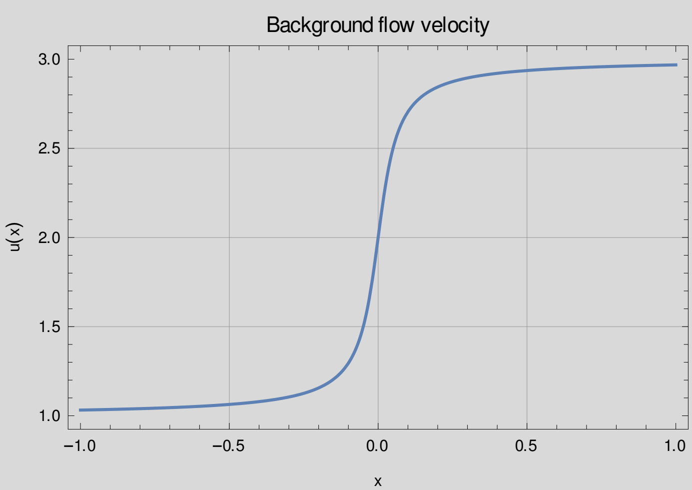
Viscoelastic flow in a convergent channel

Consider the flow of polymers in a convergent channel -1<x<1. The x-velocity of the background flow in the channel is assumed to be known:

In[ ]:=

```
u[x_] = 2/π ArcTan[20 x] + 2;
Plot[u[x], {x, -1, 1}, Axes -> False, Frame → True, FrameLabel → {"x", "u(x)"},
 GridLines → Automatic, PlotLabel → "Background flow velocity"]
```

Out[ ]=



Assuming a dilute aqueous solution, the polymers respond to changes in the liquid velocity by changing their conformation. This process is used, for instance, to stretch DNA molecules prior to identifying base pairs. The simplest model for the complicated internal conformation is a vector Q from one end of the molecule to the other end, the so-called dumbbell model.

We are interested in extracting the probability distribution functions for Q from data. In practice, this would be obtained by measurements. Here, however, we generate synthetic data from a stochastic process that models the stretching of the polymer by shear within the flow according to the stochastic differential equation

$dQ = (\eta\, u'(x)\, Q - k\, Q)\, dt + \sigma\, dW$

with η,k,σ, model constants that correspond to viscosity, dumbbell stiffness, temperature, respectively. The dW term is a Wiener process modeling the Brownian motion of the dumbbell ends. Note that Q=0 is taken to signify that the “spring” between the two ends of the dumbbell is at its equilibrium, zero-force length

Thermal effects

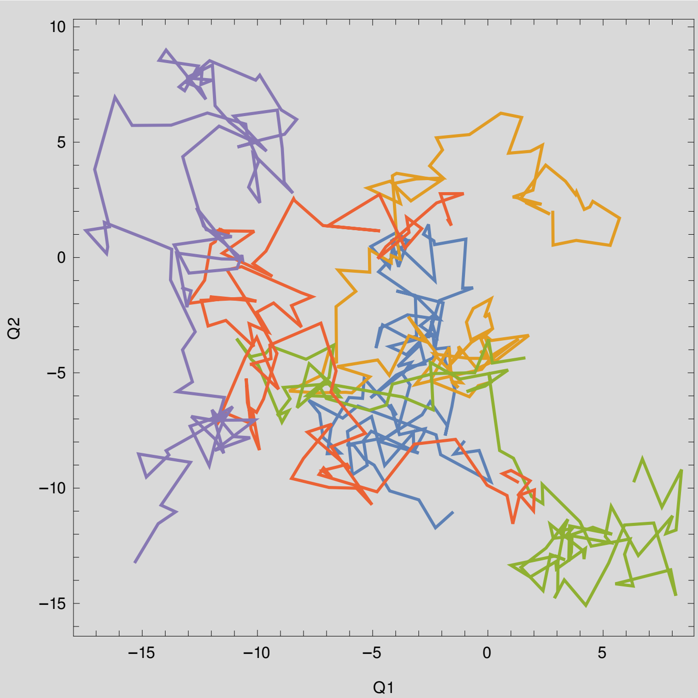
Taking η=0 leads to a model that is only influenced by temperature $dQ = -k\, Q\, dt + \sigma\, dW$, in which thermal effects force the dumbbell away from its Q=0 equilibrium length (expressed through the mean value μ=0). This is known as an Ornstein-Uhlenbeck process. Generate the synthetic data for 5 instances (i.e., 5 polymers) over 100 time steps of 0.01s.

In[ ]:=
```
η = 0; k = 1; σ = 10; μ = 0; x = 0; nPoly = 5;
proc = OrnsteinUhlenbeckProcess[μ, σ, η u'[x] + k];
db = Table[RandomFunction[proc, {0, 1, 0.01}, 2][["ValueList"]]ᵀ, {nPoly}];
```

Represent the time-evolution of the two components of Q=(Q1,Q2)

In[ ]:=
```
ListPlot[db, Joined → True, AspectRatio → Automatic,
 Frame → True, Axes → False, FrameLabel → {"Q1", "Q2"}]
```

Out[ ]=

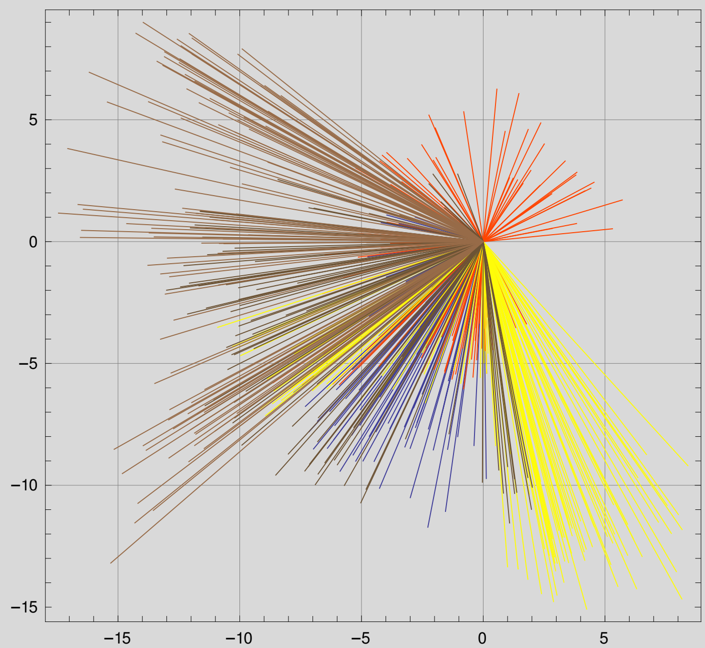


Also construct a direct representation of the evolution of the dumbbell vector itself

In[ ]:=
```
For[i = 1; dbVec = {}, i ≤ nPoly, i++,
  AppendTo[dbVec, Graphics[{ColorData[i, "ColorList"][[i]], Map[Line[{{0, 0}, #}] & , db[[i]]]}]]];
```

In[ ]:=
```
Show[dbVec, Frame → True, GridLines → Automatic]
```

Out[ ]=



The above can be interpreted as depicting the evolution over 100 time increments of the end-to-end (dumbbell) vector for nPoly=5 different polymers.

Flow effects

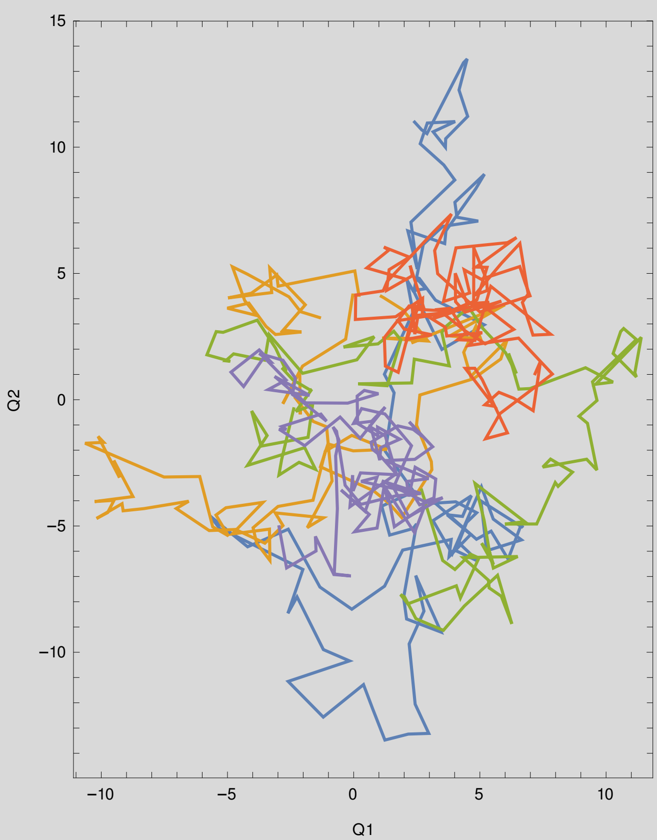
Now turn on viscosity, by setting η>0

In[ ]:=
```
η = .1; k = 1.; σ = 10.; μ = 0; x = 0; nPoly = 5;
proc = OrnsteinUhlenbeckProcess[μ, σ, η u'[x] + k];
db = Table[RandomFunction[proc, {0, 1, 0.01}, 2]["ValueList"]ᵀ, {nPoly}];
```

Represent the time-evolution of the two components of Q=(Q1,Q2)

In[ ]:=
```
ListPlot[db, Joined → True, AspectRatio → Automatic,
  Frame → True, Axes → False, FrameLabel → {"Q1", "Q2"}]
```

Out[ ]=



Karhunen-Loeve, SVD

Homework questions

Question 1

Compute the Fisher metric for univariate Gaussian distributions.

Question 2

Formulate and solve the geodesic equation for univariate Gaussian distributions. Plot the geodesics in the (μ,σ)plane, the Poincare hyperbolic plane.

Question 3

Plot the trajectory of the Gaussian PDFs for Qx,Qy obtained by the SVD in the Poincare hyperbolic plane along the channel flow for viscosities η=10^(-k), k=1,2,3,4,5.

Question 4

Arbitrarily choose one of the viscosities. Recompute the Qx,Qy PDFs for a larger number of sample paths, e.g., nPoly=250 instead of nPoly=50. Plot the trajectory in the Poincare plane of the PDFs for both the nPoly=250 and the nPoly=50 cases.

Question 5

Presumably the means of Qx, Qy obtained for the larger number of samples (nPoly=250) is more accurate. Correct the trajectory for the nPoly=50 case by geodesic transport onto the means obtained from the nPoly=250 case. Plot the initial and corrected trajectories in the Poincare plane.