1 Overview of fold continuation demos

In all demos we have a two-parameter plane of (x_1, x_2), on which measurements are taken with mean $\bar{y}$ and error with variance ν, of the form

$$y(x) \sim \bar{y}(x) + N(0, \nu(x)).$$

We want to track a curve in the (x_1, x_2)-plane of the form $f(y(x), \partial_x y(x), x) = 0$. For example, take an ODE with cusp normal form

$$\dot{u} = a + bu - u^3 \tag{1}$$

where we apply control through a

$$a(u) = k[u_{ref} - u] \quad \text{with, e.g., } k = 3. \tag{2}$$

Then, $(x_1, x_2) = (b, u_{ref})$, and for every input (x_1, x_2) we could measure

$$y(x_1, x_2) = \lim_{t\to\infty} u(t),$$

which would be the equilibrium output of (1)–(2). The fold curve is then defined by $\partial a/\partial u_{ref} = 0$, that is, $0 = k[1 - \partial_2 y(x)]$, such that we have to track the root curve of

$$0 = \partial_2 y(x_1, x_2) - 1. \tag{3}$$

1.1 Interpolation/linear regression

Suppose we already made a set of measurements in a set of points :

$$[Y, V] := (\bar{y}_1, \nu_1), \ldots, (\bar{y}_N, \nu_N) \quad \text{in} \quad X := \begin{bmatrix} x_{1,1} \\ x_{2,1} \end{bmatrix}, \ldots, \begin{bmatrix} x_{1,N} \\ x_{2,N} \end{bmatrix}$$

with means $\bar{y}_j$ and variances ν_j. Y and V are the vectors of measured means and variances of y in the points stored $2 \times N$ array X. We use some meshless linear interpolation/regression method[1] that generates functions $y(x)$ for means and $\nu(x)$ for uncertainties for all $x \in \mathbb{R}^2$. The results y and ν depend on the measurements, so lets write

$$\begin{aligned} y(x) &= y_g(x; Y, V, X) \quad \text{(expected value according to regression estimate)} \\ \nu(x) &= \nu_g(x; Y, V, X) \quad \text{(variance/uncertainty according to regression estimate).} \end{aligned}$$

Linear regression usually means, y_g is linear in Y. The functions y_g and ν_g are smooth in x such that the interpolation defines us a trackable curve via the equation

$$f(y_g(x; Y, V, X), \partial_x y_g(x; Y, V, X), x) = 0.$$

[1] I use a Gaussian process.

Specifically, for the cusp problem, this is given by solving the equation

$$0 = \frac{\partial}{\partial x_2} y_g(x_1, x_2; Y, V, X) - 1.$$

The variance v_g and some chain rule would also define a (linearised) uncertainty of the implicitly defined curve, but the simple procedure below does not rely on this. Since the regression function y_g is analytically known for given (Y, V, X), its derivatives can be computed analytically or by using finite differences such that standard Newton iterations and continuation methods can be applied to any root problem using y_g with given (Y, V, X).

1.2 Minimal continuation

Given

- Sets of measurements (Y, V) in points X, with corresponding regression functions $y_g(\cdot; Y, V, X)$ and $v_g(\cdot; Y, V, X)$.
- Point $x_c \in \mathbb{R}^2$ on the implicitly defined curve

$$f(y_g(x_c; Y, V, X), \partial_x y_g(x_c; Y, V, X), x_c) = 0.$$

- Tangent $x_{tan} \in \mathbb{R}^2$ of implicitly defined curve at x_c (of unit length):

$$\frac{\partial}{\partial x} \left[f(y_g(x; Y, V, X), \partial_x y_g(x; Y, V, X), x) \right] |_{x=x_c} x_{tan} = 0, \quad \|x_{tan}\| = 1.$$

Sensitivity We define the sensitivity of x_c with respect to a new measurement at a point x_{new}, $e(x_{new}$. Assume that we make a measurement at x_{new} that is off the expected value by one estimated standard deviation according to the regression (y_g, v_g). When including this measurement into our regression, how much would the root x_c change? Or, easier, how much would the residual of f at x_c change? More precisely,

$$\begin{aligned} y_{new} &:= y_g(x_{new}; Y, X, V) + \sqrt{v_g(x_{new}; Y, X, V)}, \\ v_{new} &:= \text{mean } V \quad \text{(rough guess)}, \qquad\qquad \text{then} \\ e(x_{new}) &:= |f(y_g(x_c; m), \partial_x y_g(x_c; m), x_c)| \quad \text{where} \quad m = [Y, y_{new}], [V, v_{new}], [X, x_{new}]. \end{aligned} \tag{4}$$

The function $e : \mathbb{R}^2 \mapsto [0, \infty]$ has a maximum point x_{max} with maximum value e_{max}. A measurement at this point will have the biggest influence on the current estimate of the root, according to our current belief in the form of regression (y_g, v_g).

Steps

1. **Improve** the accuracy of x_c with additional measurements.
   a) Set $e_{cur} = e_{max}$.
   b) Repeat until at most n_{max} points are added, or until $e_{cur} < \texttt{tolerance}$:
      i. Take measurement $(\bar{y}_{new}, v_{new})$ at x_{new}, where x_{new} is the x maximising e given in (4).
      ii. Incorporate, new measurement into regression function:

$$y(x) = y_g(x; [Y, \bar{y}_{new}], [V, v_{new}], [X, x_{new}])$$
$$v(x) = v_g(x; [Y, \bar{y}_{new}], [V, v_{new}], [X, x_{new}].$$

      and consider for following steps sensitivity e using the updated y_g and v_g (hence, x_{max} will likely be different in the next step).

2. **Correct and continue** x_c using the new interpolation (y_g, v_g) (which has taken into account the data points, added in the improvement step) according to

   $0 = f(y_g(x_{c,new}; m), \partial_x y_g(x_{c,new}; m), x_{c,new})$ where $m = (Y, V, X)$, includes new points,
   $0 = x_{tan}^T[x_{c,new} - x_c] - \texttt{stepsize}$.

3. **Reduce** interpolation. Remove any data points that do not influence the residual of f at the current new x_c. For example, sort the points in X in decreasing distance from x_c and remove points (and measurements) from $m = (Y, V, X)$ as long as

$$|x_{c,reduced} - x_c| < \texttt{threshold}, \quad \text{where}$$
$$0 = f(y_g(x_{c,reduced}; m_{reduced}), \partial_x y_g(x_{c,reduced}; m_{reduced}), x_{c,reduced})$$

   but keep, for example, at least `points2keep` points.

4. Move back to Improvement step for next step of continuation.

Remarks

- For a given interpolation (y_g, v_g) and point x_c ,finding the point x to which the sensitivity e defined in (4) is maximal, is a nonlinear optimisaton problem. It can be solved without running experiments, but it is still overkill, since the estimates are not that reliable. Instead, I pick an approximately equidistributed number (say $3n_{max}$) within a radius `checkradius` of x_c. Then I find the point which maximises e among them.

  There are also many alternatives to defining sensitivity. One can derive a variance v of x_c (the root of f) by linearising f in x_c. Then one could check for which added measurement point x this quantity v is reduced the most, assuming the new measurement $\bar{y}$ in x equals $y_g(x; Y, V, X)$ (x_c itself will not change if measurement $\bar{y}$ equals regression value $y_g(x; Y, V, X)$).

- During the improvement step, one has to run the experiment several times at a sequence of points. One can afford to have this run to be of the form for a function call (in matlab notation)

  ```
  [x,ybar,yvar,data]=RunExperiment(x_requested,data)
  ```

  where the argument `x_requested` is the set of experimental inputs that one requests to be used, but the additional output `x` is the set of inputs in which it was actually run. After the run, the new measurement $\bar{y} =$ `ybar`, $v =$ `yvar` at point `x` is included into the set of regression points (Y, V, X). Since the list of points in which measurements are to be made are only optimising approximate sesnitivity, using another point instead is not a problem.

  This permits the continuation to keep the problem of restricting to experimentally feasible points entirely inside the call to the experiment.

- It is sensible to set the tolerance for the improvement step to, for example,

  $$\texttt{tolerance}=e_{\max}/2$$

  or so (where $e_{\max}$ was the maximal sensitivity before improvement). The actual values of the sensitivity may depend on regression parameters. For example, the Gaussian provess, which I used, has a parameter `cvm` that determines how "bendy" the interpolation is (and, thus, how closely it follows measurements, smaller `cvm`=bendier). So, increasing `cvm` automatically decreases the sensitivity.

  Also, it may not make sense to set $n_{\max}$ large to insist that sensitivity drops by a certain amount. It is more economic to keep $n_{\max}$ moderate and move on even if e is not below `tolerance`.

  > The quantity $e_{\max}$ is not an error estimate, but rather an estimate, how much additional information a new measurement will give. Hence, one doesn't stop, once a certain accuracy is achieved, but rather, when the benefit of additional measurements has diminished.

- The list of points x_c produced during the continuation is not very accurate. After the continuation, one assembles a regression (y_g, v_g) over all measurements (Y, V, X) and performs a standard continuation on this regression function. This has much lower uncertainty v in its root curve of f since one is always interpolating measurements, while during the continuation x_c is at the boundary of the set of measurement points.

- The whole approach can be generalised to $x \in \mathbb{R}^n$ with $n > 2$, $y \in \mathbb{R}^m$ and $f \in \mathbb{R}^{n-1}$. The variance in the interpolation/regression the becomes an $m \times m$ covariance matrix. The functions for Gaussian process that I use assume that the covariance matrix for y is diagonal (storing only v_k for y_k).

  The advantage of regressing over many points to accommodate for noise diminishes for higher dimensions though.