Homotopy Continuation

Max Buot (CMU)
&
Donald Richards (PSU)

Computational Algorithms for Astrostatistics
Summer School in Statistics for Astronomers & Physicists

16 June 2005

What is Homotopy Continuation?

An approach to solving a system of equations, $\mathcal{S}$, by tracking the solutions of "nearby" systems of equations.

To solve $\mathcal{S}$:

- Introduce an index t, $0 \leq t \leq 1$
- For each t, construct a new system, $\mathcal{S}_t$.
- Starting with $t = 0$ and increasing t by a small step size, we solve each $\mathcal{S}_t$ in turn until we arrive at $t = 1$.
- $\mathcal{S}_0$ should be easy to solve and $\mathcal{S}_1$ is identical to $\mathcal{S}$, the original system.

A Bivariate Example

$$
\begin{aligned}
f_1(x,y) &= 3x^3 - 4x^2y + y^2 + 2y^3 \\
f_2(x,y) &= -6x^3 + 2xy - 5y^3
\end{aligned}
$$

- Problem: Solve the system of equations

$$f_1(x,y) = 0, \quad f_2(x,y) = 0$$

- How many (complex) roots are there?
- Every solution is of the form $x = y = 0$ or $x, y \neq 0$.
- **Algebraic Geometry**: The study of roots of systems of polynomial equations.
- **Bezout's Theorem**: The number of non-zero roots is <u>at most</u> $deg(f_1) \cdot deg(f_2) = 3 \times 3 = 9$.
- **Bernstein's Theorem**: The number of non-zero roots is <u>exactly</u> 4.

The HC Approach

- Introduce an index parameter t, $0 \leq t \leq 1$
- Construct a system $\mathcal{S}_t$: Let

$$g_1(x,y) = x^3 - 1, \quad g_2(x,y) = y^3 - 8$$

- Define

$$\begin{aligned} h_1(x,y;t) &= tf_1(x,y) + (1-t)g_1(x,y) \\ h_2(x,y;t) &= tf_2(x,y) + (1-t)g_2(x,y) \end{aligned}$$

- $\mathcal{S}_t$: $\quad h_1(x,y;t) = 0, \quad h_2(x,y;t) = 0$
- $\mathcal{S}_0$: $\quad g_1(x,y) = 0, \quad g_2(x,y) = 0$
- $\mathcal{S}_0$ is easily solved.
- $\mathcal{S}_1$: $\quad f_1(x,y) = 0, \quad f_2(x,y) = 0$
- $\mathcal{S}_1$ is the original system.

The Bivariate Example, cont.

$$\begin{aligned} f_1(x,y) &= 3x^3 - 4x^2y + y^2 + 2y^3 \\ f_2(x,y) &= -6x^3 + 2xy - 5y^3 \end{aligned}$$

- Problem: Solve the equations

$$f_1(x,y) = 0, \quad f_2(x,y) = 0$$

- By Bernstein's theorem, there are *exactly* 4 non-zero roots.
- Verschelde (1999), "Algorithm 795: PHCPack: A general-purpose solver for polynomial systems by homotopy continuation," *ACM Trans. Math. Software*, 25 (1999), 251–276. www2.math.uic.edu/∼jan/PHCpack/phcpack.html
- Let us now use PHCPack to find the roots. (Demo)

PHCpack Results

```
solution 2 :
x :    -3.16950027102798E-02   1.81213765826737E-01
y :    -1.10462903286809E-01  -2.04565439823804E-01
solution 7 :
x :    -3.16950027102798E-02  -1.81213765826737E-01
y :    -1.10462903286809E-01   2.04565439823804E-01
solution 8 :
x :    4.11875744374350E-01  -1.20485325502200E-01
y :    1.67490014536420E-01  -3.32145613080015E-01
solution 9 :
x :    4.11875744374350E-01   1.20485325502200E-01
y :    1.67490014536420E-01   3.32145613080015E-01
```

Advantages of HC

- Applicable to optimization problems which can be reduced to systems of equations.
- Are numerical stable.
- Can be globally convergent, rather than only locally convergent.
- Can locate multiple solutions to $F(\mathbf{x}) = 0$.
- Can provide insight into properties of the solutions.
- Can be modified to located real solutions only.
- Newton-type methods can be used to find solutions of the deformed system at various $t-$increments.

History and References

- Poincaré (1881-1886), Klein (1892-1883)
- Bernstein (1910), Leray and Schauder (1934)
- Zangwill and Garcia (1981), *Pathways to Solutions, Fixed Points, and Equilibria*, Prentice-Hall.
- Morgan (1987), *Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems*, Prentice Hall.
- Allgower and Georg (1990), *Numerical Continuation Methods: An Introduction*, Springer.

The General Case

- Variables $\mathbf{x} = (x_1, \ldots, x_m)$
- Polynomials $f_1(\mathbf{x}), \ldots, f_m(\mathbf{x})$
- Problem: Solve the system of equations

$$f_i(\mathbf{x}) = 0, \qquad i = 1, \ldots, m.$$

- Let

$$F(\mathbf{x}) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix}, \qquad G(\mathbf{x}) = \begin{pmatrix} g_1(\mathbf{x}) \\ g_2(\mathbf{x}) \\ \vdots \\ g_m(\mathbf{x}) \end{pmatrix}$$

- The solutions to g_i are known.

General Case, cont.

- Construct the *homotopy map*

$$H(\mathbf{x}; t) = tF(\mathbf{x}) + (1-t)G(\mathbf{x})$$

- To solve $F(\mathbf{x}) = 0$, we track the solution paths of $H(\mathbf{x}; t) = 0$ from $t = 0$ to $t = 1$.

⋆ *Homotopy continuation*: The process of path-tracking.

General Case, etc.

- There are many other homotopies (or *deformations*). The *global homotopy*:

$$H(\mathbf{x};t) = F(\mathbf{x}) - (1-t)F(\mathbf{x}_0)$$

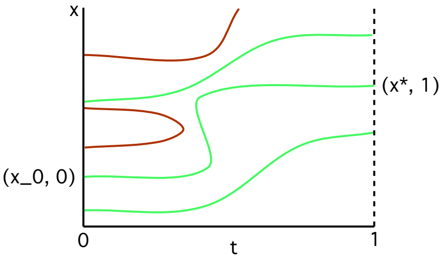
  where x_0 is a starting point.

- A general homotopy:

$$H(\mathbf{x};t) = \phi(t)F(\mathbf{x}) + \psi(t)G(\mathbf{x})$$

  where:
  $\phi(0) = \psi(1) = 0$, $\phi(1) = \psi(0) = 1$, and
  $\phi(t) > 0$ and $\psi(t) > 0$ whenever $0 < t < 1$

General Case, etc.



- $\mathbf{x}^*$: A solution of the system $F(\mathbf{x}) = 0$
- $\mathbf{x}_0$ is a starting point for the homotopy continuation process
- The *zero set* of the homotopy:

$$\mathcal{H}_0 = \{\mathbf{x} : H(\mathbf{x}; t) = 0 \text{ for some } t \in [0, 1]\}$$

- In homotopy continuation, we need to find a curve c in $\mathcal{H}_0$ which runs from $(x_0, 0)$ to $(x^*, 1)$

Questions

1. When does such a curve exist and is smooth?
2. Will its length be finite?

Let $h_1, \ldots, h_m$ be the functions comprising $H(\mathbf{x}; t)$. Construct the *Jacobian matrix*,

$$J_{\mathcal{H}}(\mathbf{x}; t) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_m} & \frac{\partial h_1}{\partial t} \\ \frac{\partial h_2}{\partial x_1} & \cdots & \frac{\partial h_2}{\partial x_m} & \frac{\partial h_2}{\partial t} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial h_m}{\partial x_1} & \cdots & \frac{\partial h_m}{\partial x_m} & \frac{\partial h_m}{\partial t} \end{pmatrix}$$

The Existence and Solution of Paths

The existence of c is a consequence of the *Implicit Function Theorem*:

1. If $J_{\mathcal{H}}(\mathbf{x};t)$ has full rank n at $(\mathbf{x}_0;0)$ then there exists (at least locally) a smooth curve c which starts at $(x_0;0)$.
   If $J_{\mathcal{H}}(\mathbf{x};t)$ has full rank n at all points in $\mathcal{H}$ then c is "diffeomorphic" to a circle or line.
2. For the curve c to have finite length, we impose boundary conditions to prevent the curve from going to infinity before going to $(\mathbf{x}^*;1)$ or from returning to $(\mathbf{x}_0;0)$.

LF Data

Globular cluster luminosity functions in the Galaxy

A table from Secker 1992, AJ 104, 1472:

Milky Way LF Example, Revisited

In the "maximum likelihood estimation" lecture of last week, we discussed a t-distribution model (Secker 1992, AJ 104, 1472) for LF data:

$$g(x;\mu,\sigma,\delta) = \frac{\Gamma(\frac{\delta+1}{2})}{\sqrt{\pi\delta}\,\sigma\,\Gamma(\frac{\delta}{2})}\left[1 + \frac{(x-\mu)^2}{\delta\sigma^2}\right]^{-\frac{\delta+1}{2}}$$

$-\infty < \mu < \infty, \sigma > 0, \delta > 0$

- Given a random sample $x_1, \ldots, x_n$ of LF values, we wish to calculate the MLEs of μ, σ, and δ.
- There are no explicit formulas for these MLEs; we must use numerical methods to maximize the likelihood function.

Milky Way LF Example, cont.

- For simplicity, we assume that $\delta = 3.55$, the value of the MLE given by Secker for the Milky Way data.
- We now have a two-parameter likelihood function:

$$\begin{aligned} L(\mu, \sigma^2) &= g(x_1; \mu, \sigma^2) \cdots g(x_n; \mu, \sigma^2) \\ &= \prod_{i=1}^{n} \frac{\Gamma(\frac{\delta+1}{2})}{\sqrt{\pi\delta}\, \sigma\, \Gamma(\frac{\delta}{2})} \left[1 + \frac{(x_i - \mu)^2}{\delta\sigma^2}\right]^{-\frac{\delta+1}{2}} \end{aligned}$$

$$\begin{aligned} \ln L = n \ln \frac{\Gamma(\frac{\delta+1}{2})}{\sqrt{\pi\delta}\, \Gamma(\frac{\delta}{2})} - \frac{1}{2} \ln \sigma^2 \\ - \frac{\delta+1}{2} \sum_{i=1}^{n} \ln \left[1 + \frac{(x_i - \mu)^2}{\delta\sigma^2}\right]^{-\frac{\delta+1}{2}} \end{aligned}$$

Milky Way LF Example, cont.

- Maximum likelihood equations:

$$\frac{\partial \ln L}{\partial \mu} = 0, \quad \frac{\partial \ln L}{\partial (\sigma^2)} = 0.$$

- Reduce each equation to polynomial form.
- The corresponding *score equations* are:

$$\sum_{i=1}^{n} (x_i - \mu) \prod_{\substack{j=1 \\ j \neq i}}^{n} [3.55\sigma^2 + (x_i - \mu)^2] = 0$$

and

$$\sum_{i=1}^{n} (x_i - \mu)^2 \prod_{\substack{j=1 \\ j \neq i}}^{n} [3.55\sigma^2 + (x_i - \mu)^2] - \frac{n}{4.55} \prod_{j=1}^{n} [3.55\sigma^2 + (x_i - \mu)^2] = 0$$

respectively.

- We have two equations in two variables, μ and σ^2.

PHCpack: Milky Way LF Example

- For the sake of simplicity, let $n = 6$
- Data points: $-3.88, -5.52, -7.03, -7.77, -8.29, -9.25$
- PHCpack finds two admissible solutions (among 26 total)

- ```
  solution 66 :
  a :  -3.88384638288804E+00   6.57457755788503E-26
  b :   6.58244300380849E-04   4.46858148371494E-26
  solution 72 :
  a :  -7.21052837109880E+00   8.36412639400878E-25
  b :   2.24221225822324E+00   3.84173364494972E-24
  ```

- The solution at the end of path 72 is the MLE.
  ```

Return to bivariate example on slide 3.

The fact that the system has exactly 4 non-zero roots is a consequence of Bernstein's Theorem (see Cox, et. al., 2000).

To apply Bernstein's Theorem,

(1) We find the monomials which appear in each polynomial. For example, the monomials in f_1 are $(3,0),(0,2),(0,3),(2,1)$.

(2) We calculate the convex hulls of each set of monomials. Denote these convex hulls by $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively.

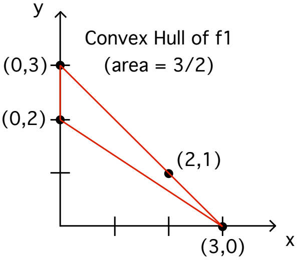
(3) We calculate the "Minkowski Sum" of these two convex hulls, denoted $\mathcal{C}_1 + \mathcal{C}_2$.

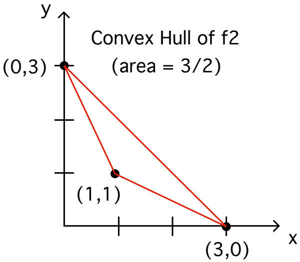
(4) Finally, calculate the areas of the convex hulls $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_1 + \mathcal{C}_2$. Denote these areas by A_1, A_2, and A_{12}, respectively.

(5) Finally, Bernstein's Theorem states that the number of non-zero roots is $A_{12} - A_1 - A_2$.

Convex Polytopes and Bernstein's Theorem

The convex hulls of f_1 and f_2 on slide 3:



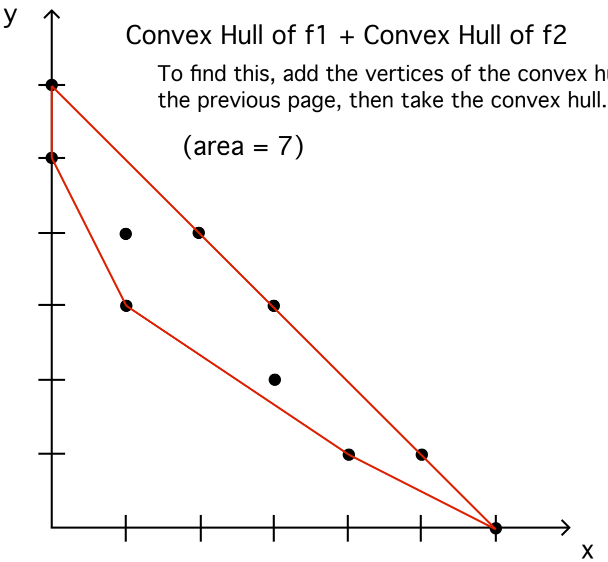


The Minkowski Sum

Convex Hull of f1 + Convex Hull of f2

To find this, add the vertices of the convex hulls on the previous page, then take the convex hull.

(area = 7)



References

- Allgower, E. and Georg, K. (1990) Numerical continuation methods: an introduction, Springer-Verlag, New York.
- Buot, M. (2003) Genetic Algorithms and Maximum Likelihood Estimation. Ph.D. Dissertation, University of Virginia.
- Buot, M. and Richards, D. (2005) Counting and locating the solutions of polynomial systems of maximum likelihood equations, I. Journal of Symbolic Computation, to appear.
- Chow, J., Udpa, L., and Udpa, S.S. (1991) Homotopy continuation method for neural networks. *IEEE Proc ISCAS Singapore*.
- Coetzee F. M. and Stonick V. L. (1996) On a natural homotopy between linear and nonlinear single layer preceptron networks. *IEEE Transactions on Neural Networks*, 7, 307-317.
- Coetzee, F. M. (1995) Homotopy approaches for the anaylsis and solution of neural network and other nonlinear systems of equations. Ph.D. Dissertation, Carnegie Mellon University.
- Cox, D. A., Little, J., and O'Shea, D. Using algebraic geometry, Springer.
- Chu, M. T. (1984) A simple application of the homotopy method to symmetric eigenvalue problems. *Linear Algebra and its Applications*, 59, 85-90.
- Chu, M. T. (1988) A note on the homotopy method for linear algebraic eigenvalue problems. *Linear Algebra and its Applications*, 105, 225-236.
- Miller, W. D. (1992) Homotopy analysis of recurrent neural nets. *Digital Signal Processing*, 2, 33-38.
- Morgan, A. (1987) Solving polynomial systems using continuation for engineering and scientific problems, Prentice Hall, New Jersey.