NON-QUADRATIC METHODOLOGIES FOR PROCESS OPTIMIZATION

Angelo Lucia* and Peter A. DiMaggio
University of Rhode Island
Kingston, RI 02881-0805

Abstract

In a recent paper, Lucia et al. (2003a) introduced a new class of non-quadratic functions for modeling the large-scale geometry of general objective functions called generalized exponential funnel functions. This paper provides additional evidence to show that exponential funnels and other non-quadratic functions are often better models for objective function surfaces and can result in significant improvements in numerical performance in both local and global optimization. Two simple chemical engineering examples are presented to support these claims.

Keywords

Non-quadratic functions, Exponential funnel functions, Asymptotes, Cusps.

Introduction

In optimization, we have long subscribed to the notion that one size fits all. By this we mean that most, not all, well-established numerical methods in widespread use in optimization are based on quadratic Taylor series expansions of the objective function and/or constraints. This has given rise to a number of algorithms that all fit under the umbrella of Newton-based methods (e.g., Newton's method, quasi-Newton methods, successive quadratic programming, and so on). Although there are a small number of methods that are based on non-quadratic approximations like the Jacobson-Oksman algorithm (1972) for homogeneous functions, the conic approximation and collinear scaling methods of Davidon (1980) and the rational function approach of Banerjee et al. (1985), these methods are rarely used or cited. It seems that we have long accepted the idea that quadratic approximation is best, despite the fact that it is easily illustrated that this is not true. For example, the Lennard-Jones 6-12 potential is clearly non-quadratic and Newton's method often performs poorly over a large portion of the feasible region on this simple function – unless some type of 'fix' is used to correct or override the Newton step. There are other physical examples like models of phase transitions and phase equilibrium problems modeled by equations of state that contain asymptotes, poles, cusps and corners within the feasible region that are also non-quadratic and exhibit what we call retrograde curvature in the neighborhood of a non-differentiable extremum. This retrograde curvature coupled with discontinuity or non-differentiability is always non-quadratic and can lead to convergence difficulties for Newton-like methods.

The main objective of this paper is to study the use of non-quadratic approximations of objective function surfaces in local/and global optimization. Two different non-quadratic functions for local or global optimization are presented to show that non-quadratic models can provide theoretical and computational advantages when used in the correct context.

* To whom all correspondence should be address

Non-Quadratic Function Approximation

The reasons for the widespread use of quadratic approximations in optimization are that they provide a simple and unambiguous way of calculating iterative estimates of an optimal solution and result in fast convergence. Any positive definite quadratic function has a unique minimum that can be calculated by solving a linear system. In an iterative context, quadratic function approximations usually give superlinear or quadratic convergence.

In this section, we show that non-quadratic functions like the generalized exponential funnel functions of Lucia et al. (2003a) and the generalized cusp functions introduced in this paper share these basic mathematical properties and, at the same time, offer additional theoretical and computational advantages. That is, funnels and cusps also have unique minima (or maxima) that can be easily calculated by solving a linear system of equations and used as iterative estimates of the solution. However, they are non-convex functions that often provide a better model of the objective function than a quadratic model over a larger portion of the feasible region and frequently require fewer function and gradient evaluations than Newton-based methods to find solutions. When used correctly, funnels and cusps are also unaffected by singularities. On balance, however, funnel functions require additional storage (at least two function, gradient and Hessian values) and additional work per iteration (the solution of a cubic

polynomial) to build an approximation of the objective function.

Generalized Exponential Funnel Functions

Lucia et al. (2003a) introduced generalized exponential funnel functions to approximate the large-scale geometry of 'rough' or 'rugged' objective function surfaces in global optimization – like the free energy surface in protein folding problems. The basic idea here is to build an approximation of the form

$$F(z) = F_0 - \Gamma e^{-q(z)} \quad (1)$$

where $q(z) = \frac{1}{2}z^T Az + b^T z + c$, and where $\Gamma > 0$, F_0 and c are scalar parameters, b is an n-dimensional vector and A is an n x n symmetric matrix. From the gradient of the funnel function in Eq. (1), it is easy to show that the unique minimum of the funnel is $y = A^{-1}b$. To find A and b, a scaling parameter, $\gamma = \max(\gamma_k, \gamma_{k+1}) > 0$ must be determined by using the Routh criterion (Routh, 1905) and solving a cubic equation that arises from matching function, gradient and Hessian matrix information of any arbitrary twice continuously differentiable objective function, say f(z), g(z) and h(z), with function, gradient and Hessian matrix values of the funnel function, F(z), G(z) and H(z), at two distinct points. This results in

$$A = [\gamma h + gg^T]/\gamma^2 \quad (2)$$

and

$$b = g/\gamma - Az \quad (3)$$

From γ, A and b, values of $\Gamma > 0$, F_0 and c can be computed directly. Moreover, minimum eigenvalue estimation methods can be used to ensure that A is positive definite. This, in turn, guarantees that the funnel has the right shape and its minimum is unique. See Lucia et al. (2003a) for details.

In our opinion, generalized funnel functions offer the following advantages over quadratic approximations (or Newton methods). They are non-convex and provide better approximations to arbitrary objective functions over a wider range of the variables. They have unique minima that can be calculated by solving a linear system of equations – just as in Newton's method. They are virtually unaffected by singular points because of the freedom built into these funnel approximations, largely through the matrix A. They are self-correcting in the face of convexity changes because they retain function, gradient and Hessian matrix values at two or more points. Finally, they have quadratic convergence. Thus for roughly the same amount of work per iteration (i.e., the solution of a linear system) and the retention of a small amount of extra information, generalized exponential funnels provide a self-correcting extension of Newton's method that, to date, has performed well in practice.

Generalized Cusp Functions

Many problems in science and engineering – problems in cosmology, physiology, oceanography, fluid dynamics and others – contain isolated points or manifolds within the feasible region where the governing model is discontinuous and/or non-differentiable; elsewhere the model is smooth. For optimization problems with discontinuous and/or non-differentiable points we suggest the use of generalized cusp functions of the form

$$C(z) = C_0 + [(z - z_c)^T A(z - z_c)]^{1/m} \quad (4)$$

where C_0 is a constant, z_c is the location of the cusp, A is a symmetric, positive definite matrix of order n and $m > 2$ is an integer. Note that C(z) is a bounded function that is only non-differentiable at its minimum - the cusp.

For n = 1, it follows that

$$C'(z) = [2A^{1/m}(z - z_c)^{(2 - m)/m}]/m \quad (5)$$

for z unequal to z_c and

$$C(z)'' = 2(2\text{-}m)A^{1/m}(z - z_c)^{(2 - 2m)/m}]/m^2$$

$$= [(2\text{-}m)/m]C'(z)/(z - z_c) \quad (6)$$

Clearly Eq. (6) implies

$$(z - z_c) = [(2\text{-}m)/m]\, C'(z)/C''(z) \quad (7)$$

or that when the objective function is truly given by Eq. (4) the cusp can be computed in a single iteration by simply scaling the Newton step by (2 – m)/m. Nothing could be simpler! Note that C'(z) and C''(z) approach infinity as z approaches z_c, and that this simple analysis also shows full Newton steps will not efficiently locate a cusp because C(z) is strongly non-quadratic in the neighborhood of a cusp. Full Newton steps are too large and will either zigzag their way down both sides of the cusp or move outside of the feasible region. Thus we select m as the largest whole number in the integer sequence {3, 4, … } that minimizes C(z) and also maintains feasibility in the neighborhood of the cusp.

For n > 1, the mathematics is not 'clean'. However, one straightforward way to extend this simple analysis to multivariable cusps, corners, asymptotes and so on is by direct analogy. That is, iterative estimates of the cusp should be calculated using

$$(z - z_c) = [(2\text{-}m)/m]\, H_c^{-1}(z)G_c(z) \quad (8)$$

where $G_c(z)$ and $H_c(z)$ are the gradient and Hessian matrix of the multivariable cusp function defined in Eq. (4). Again, when C(z) is truly a multivariable cusp function the minimum or cusp point is located in a single iteration. For arbitrary objective functions we replace $G_c(z)$ and $H_c(z)$ of the cusp function with g(z) and h(z) and iterate using Eq. (8) to find the cusp point.

When To Use Non-Quadratic Approximations

We have always made the choice to use generalized exponential funnels a priori (e.g., in an outer loop in the multi-scale global optimization of objective functions with rugged terrains). See, again, Lucia et al. (2003a). However, generalized cusps can be invoked on the fly by monitoring both the Newton step and curvature along any terrain path (Lucia and Feng, 2002, 2003) to locate asymptotes, poles, cusps or corners.

Can We Really Do Better Than Newton's Method?

Yes! We present two small chemical engineering examples to show that methods based on non-quadratic objective function approximations can provide better numerical performance than Newton's method. These examples were specifically chosen because they do not have the same form as the non-quadratic approximation. Thus they challenge the robustness of the theoretical and computational framework.

Example 1: The Lennard-Jones Potential

The Lennard-Jones 6-12 potential is commonly used to model non-bonded or van der Waal's forces in molecular simulation and is given by

$$E_{LJ}(r) = 4\varepsilon[(\sigma/r)^{12} - (\sigma/r)^{6}] \qquad (9)$$

where ε is an energy parameter, σ is a distance parameter, r is an unknown separation between a pair of particles and the feasible region is $r \in (\sigma, 4]$. Although the minimum of $E_{LJ} = -\varepsilon$ at $r = 2^{1/6}\sigma$ can be computed analytically, the computational challenge presented by this problem is the strongly non-quadratic nature of the Lennard-Jones function. There is an inflection point of E_{LJ} at $r = 1.244455059\,\sigma$ and a singular local maximum at $r = 4$ that can attract Newton iterates. When direct prediction Newton's method is started with any $r > 1.244455059\,\sigma$, it fails to find the minimum of E_{LJ} because of this strong non-convexity. Although the minimum can be easily located by reversing the Newton direction so that it's a descent direction on E_{LJ}, this strategy does not always work. Besides, there is still uncertainty about the global nature of the solution.

If, instead, we start at say $r = 4.216867\,\sigma = 7$ and use the terrain method of Lucia et al. (2003b), then 109 function and gradient calls are needed - 18 for the local maximum, 53 for the inflection point, 21 for the global minimum, and 17 to cover the rest of the feasible region. The actual bounds used in these calculations were given by $r \in (\sigma, 50]$ and convergence was assumed when $\|g\| < 10^{-8}$, where g denotes the gradient of E_{LJ}.

Finally, if we again start at $r = 4.216867\,\sigma = 7$, take one Newton step, and use values of f, g and h at both the starting point and first iterate to start the funneling algorithm of Lucia et al. (2003a), the global minimum is computed in 10 function and gradient evaluations. What's important about this is that although the first Newton step moves away from the global minimum, the funneling algorithm automatically corrects this and drives all subsequent iterates to the global minimum. Table 1 gives the funnel iterates for this example.

Table 1. Funnel Iterates for Lennard-Jones Potential

Iteration	γ	r (10^{-10} m)	E_{LJ}(10^{-21} J)
1	0.00019	6.89477	-0.00265
2	0.00018	2.95359	-0.41513
3	0.29224	2.65508	-0.76379
4	0.55501	2.36894	-1.41951
5	1.14792	2.06160	-2.69636
6	2.69490	1.79376	-3.17669
7	3.17471	1.83800	-3.37518
8	3.37328	1.85979	-3.39956
9	3.39771	1.86322	-3.40000
10	3.39815	1.86329	-3.40000

Figure 1 shows a few of the iterative funnel approximations of the Lennard-Jones potential.

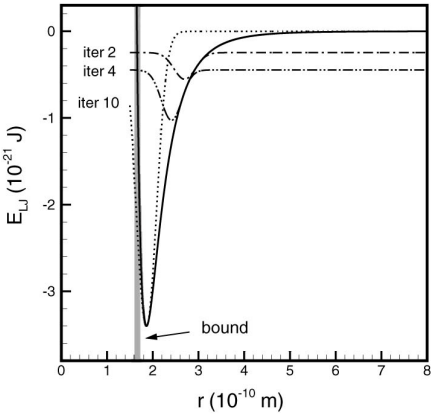

Figure 1. Funnel Approximations of LJ Potential

Example 2: Disk to Vesicle Transitions

One way to model phase transitions in soft colloids is to use the dimensionless energy proposed by Lipowsky (1992) given by

$$E(x) = (x - ba)^2 + b(1-(x/2)^2)^{1/2} \qquad (10)$$

where a and b depend on the spontaneous curvature and bending modulus of the colloidal material, x is the continuous curvature and the feasible region is $x \in [-2, 2]$. Note that $g = E'(x)$ goes to minus infinity at the boundaries because it contains the term $-(1-(x/2)^2)^{-1/2}$.

Lucia and Feng (2002) solved this problem for $a = 0.1$ and $b = 3$ using their terrain method and required 275 function and gradient calls to find the five interior stationary points and two poles of E'(x). They also remarked that roughly 200 of those function and gradient calls were due to uphill exploration on $g^T g$.

Using a similar terrain approach we re-solved this problem for two sets of parameters a and b. However, we also incorporated the cusp model given in Eq. (4) within our terrain method with a test to invoke it on the fly. This test monitors retrograde curvature and the ratio $R = \| g \| / \| \Delta_N \|$, where $\Delta_N = (z - z_k)$ is the full Newton step. If $R \geq 10^6$ and there is retrograde curvature at both z and z_k, we scale the Newton step by $\alpha = (2\text{-}m)/m$. Otherwise, the step-size is determined by the error in the Taylor series expansion from z_k to z.

We studied the performance of our terrain method with and without the on-the-fly cusp model. For a = 0.1 and b = 3 and the same starting point used by Lucia and Feng, $x_0 = 0.9$, we observed the following when no cusp model was used. For a fixed conservative uphill step-size of $\alpha = 0.1$, 265 function and gradient evaluations were required to find all of the points in Table 2 and correctly identify the global minimum at x = 2. Many of these uphill iterations (158) were used in the neighborhood of the boundaries since it is here that E(x) is far from quadratic. If a fixed aggressive step-size of $\alpha = 1$ is used the terrain method fails to find either of the cusps at x = +/-2 because uphill Newton steps cross the boundaries of the feasible region without ever detecting

Table 2. Stationary Points & Cusps of Vesicle Example

Point	x	E(x)
local min	0.489195	2.94467
singularity	-1.38561	5.00466
singularity	1.38561	3.34192
saddle	-1.89233	5.77736
saddle	1.78531	3.55837
cusp	-2.00000	5.29006
global min	2.00000	2.89000

anything. If, instead, an on-the-fly cusp model is used within our terrain method, all points in Table 2 are found in 114 function and gradient evaluations with m = 4. Moreover, the marked improvement in numerical performance is directly due to the non-quadratic cusp model. Whereas 158 function and gradient calls were needed when a quadratic model was used in the neighborhood of the cusps at x = +/- 2, when the cusp model is used this requirement is reduced to 35. Figure 2 depicts iterative behavior near the cusp. For a = 0.1 and b = 2, the global minimum is not a constrained minimum but an unconstrained minimum at x = 0.267475. Again, using the same starting point, on-the-fly cusp approximations within our terrain method provided complete reliability and needed 128 function and gradient calls to find everything. Conservative path following without the cusp approximations needed 287 function and gradient calls while aggressive path following failed to locate either cusp.

Conclusions

Two non-quadratic models for approximating general objective functions, exponential funnels and generalized

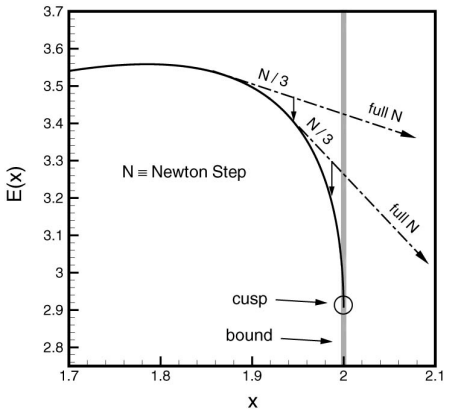

Figure 2. Iterative Behavior for m = 3 Near a Cusp

cusps, were presented. Both models are non-convex functions with unique minima and cusp functions, in particular, are capable of handling non-differentiable points within the feasible region. Numerical results for two simple chemical engineering examples were presented that show that non-quadratic approximations of general objective functions can provide reliable and efficient computations in local and/or global optimization. Copies of all numerical results in this paper are available from the authors.

Acknowledgments

The authors wish to thank the National Science Foundation for support of this work under Grant No. CTS - 0113091.

References

Banerjee, A., Adams, N., Simons, J., Shepard, R. (1985). Search for Stationary Points on Surfaces. *J. Phys. Chem., 89*, 52.

Jacobson, D.H., Oksman, W. (1972). An Algorithm That Minimizes Homogeneous Functions of N Variables in N + 2 Iterations and Rapidly Minimizes General Functions. *J. Math Anal. Appl., 38*, 535.

Davidon, W.C. (1980). Conic Approximations and Collinear Scalings for Optimizers. *SIAM J. Numer. Anal., 17*, 268.

Lipowsky, R. (1992). Budding of Membranes Induced by Intra-membrane Domains. *J. Physics II France, 2*, 1825.

Lucia, A., DiMaggio, P. A., Depa, P. (2003a). Funneling Algorithms for Multi-Scale Optimization on Rugged Terrains. *I & EC Res.*, in press.

Lucia, A., DiMaggio, P. A., Depa, P. (2003b). A Geometric Terrain Methodology for Global Optimization. *J. Global Optim.* in press.

Lucia, A., Feng, Y. (2002). Global Terrain Methods. *Comput. & Chem. Engng., 26*, 529.

Lucia, A., Feng, Y. (2003). Multivariable Terrain Methods. *AIChE J.,49*, 2553.

Routh, E.J. (1905). *Dynamics of a System of Rigid Bodies, Part II*, MacMillan, London.