Barycenter of Weight Coefficient Region of Least Weighted H^2 Norm Updating Quadratic Models for Derivative-Free Optimization

(Scan this for codes)

Pengcheng Xie (pxie@lbl.gov) and Stefan M. Wild (wild@lbl.gov)

Applied Mathematics and Computational Research Division, Lawrence Berkeley National Laboratory

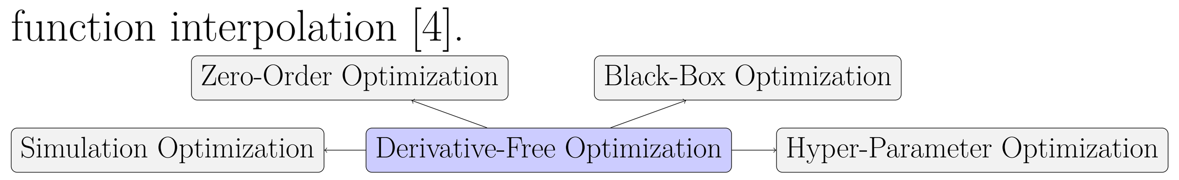
BERKELEY LAB
COMPUTING SCIENCES
Applied Mathematics & Computational Research

Abstract

Derivative-free optimization [1] problems, where objective function derivatives are unavailable, can be addressed using local quadratic models within a trust-region algorithm [5]. We propose a model updating approach when high accuracy is demanded, such as when the trust-region radius vanishes. Our approach uses the barycenter of a particular coefficient region and is shown to be advantageous in numerical results.

Introduction and Algorithm

Model-based trust-region methods iteratively select a new point for evaluation by minimizing a local model of the objective. Typical model forms include interpolation polynomials (e.g., linear interpolation, quadratic interpolation, under-determined quadratic interpolation [2, 3]), regression polynomials, and radial basis function interpolation [4].



Unconstrained derivative-free optimization problem

$$\min_{\boldsymbol{x}\in\Re^n} f(\boldsymbol{x}) \tag{1}$$

Model-based derivative-free trust-region algorithms

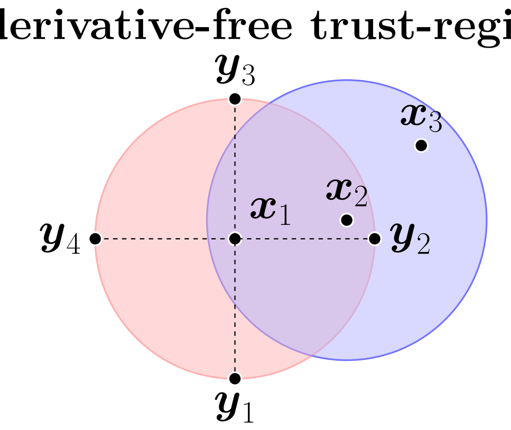


Figure 1:Demo of model-based trust-region algorithms

- **Interpolation model**
  Use a (quadratic) model Q_k to interpolate f on $\mathcal{X}_k$
- **Trust-region iteration**
  Calculate
  $\boldsymbol{x}_{k+1} \in \{\arg\min_{\boldsymbol{x}} Q_k(\boldsymbol{x}),\ \text{subject to } \|\boldsymbol{x}-\boldsymbol{x}_k\|_2 \le \Delta_k\}$
- **Model updating**
  In the k-th iteration we update the quadratic model via

$$Q_k := Q_{k-1} + Q_{\text{update}}$$

Preliminary

Definition. Let u be a function over $\Omega \subseteq \Re^n$. If u is twice differentiable on Ω and $\frac{\partial^\alpha u}{\partial \boldsymbol{x}^\alpha} \in L^2(\Omega)$ for any $|\alpha| \le 2$. We define

$$|u|_{H^0(\Omega)} = \left(\int_\Omega |u(\boldsymbol{x})|^2 d\boldsymbol{x}\right)^{\frac{1}{2}},$$

$$|u|_{H^1(\Omega)} = \left(\int_\Omega \|\nabla u(\boldsymbol{x})\|_2^2 d\boldsymbol{x}\right)^{\frac{1}{2}},$$

$$|u|_{H^2(\Omega)} = \left(\int_\Omega \|\nabla^2 u(\boldsymbol{x})\|_F^2 d\boldsymbol{x}\right)^{\frac{1}{2}},$$

and the H^2 norm of the function u via

$$\|u\|_{H^2(\Omega)} = \left(|u|_{H^0(\Omega)}^2 + |u|_{H^1(\Omega)}^2 + |u|_{H^2(\Omega)}^2\right)^{\frac{1}{2}}.$$

The Least Weighted H^2 Norm Updating Quadratic Model

Least weighted H^2 norm updating quadratic model Q_{update} [5] is obtained by solving

$$\min_{Q\in\mathcal{Q}} \sum_{i=0}^{2} C_{i+1} |Q - Q_{k-1}|_{H^i(\Omega)}^2 \quad \text{subject to } Q(\boldsymbol{y}) = f(\boldsymbol{y}),\ \boldsymbol{y} \in \mathcal{X}_k, \tag{2}$$

where $C_1, C_2, C_3 \ge 0$ are weight coefficients satisfying $C_1 + C_2 + C_3 \neq 0$ and $\mathcal{X}_k$ is the interpolation set at the k-th iteration. The solution of (2) can be obtained by solving a linear system according to the Karush–Kuhn–Tucker conditions, i.e.,

$$\text{KKT matrix: } \boldsymbol{W} \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \\ c \\ \boldsymbol{g} \end{pmatrix} = \begin{pmatrix} f(\boldsymbol{y}_1) \\ \vdots \\ f(\boldsymbol{y}_m) \\ \boldsymbol{0}_{n+1} \end{pmatrix}$$

and then obtain the model based on $(\lambda_1, \ldots, \lambda_m, c, \boldsymbol{g}^\top)^\top$.

Key Theoretical Results

Theorem. For two KKT matrices $\boldsymbol{W}$ and $\boldsymbol{W}$, $\|\boldsymbol{W}-\boldsymbol{W}^*\|_F^2$ is a function of $C_1, C_2, C_1^*, C_2^*, n, r$, and in details, it holds that

$$\begin{aligned}\|\boldsymbol{W}-\boldsymbol{W}^*\|_F^2 :=& \mathcal{D}(C_1, C_2, C_1^*, C_2^*, n, r)\\ =& \mathcal{R}_1(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m)\mathcal{P}_1(C_1, C_2, C_1^*, C_2^*, n, r)\\ &+ \mathcal{R}_2(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m)\mathcal{P}_2(C_1, C_2, C_1^*, C_2^*, n, r)\\ &+ \mathcal{R}_3(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m)\mathcal{P}_3(C_1, C_2, C_1^*, C_2^*, n, r)\\ &+ \mathcal{P}_4(C_1, C_2, C_1^*, C_2^*, n, r),\end{aligned} \tag{3}$$

where C_1, C_2 and C_1^*, C_2^* are the corresponding weight coefficients of $\boldsymbol{W}$ and $\boldsymbol{W}^*$ respectively, and r is the radius of $\Omega := \mathcal{B}_2^r(\boldsymbol{x}_0)$. The terms

$$\mathcal{R}_1(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m) = \sum_{i=1}^m \sum_{j=1}^m \left((\boldsymbol{y}_i - \boldsymbol{x}_0)^\top (\boldsymbol{y}_j - \boldsymbol{x}_0)\right)^4,$$

$$\mathcal{R}_2(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m) = \sum_{i=1}^m \sum_{j=1}^m \|\boldsymbol{y}_i - \boldsymbol{x}_0\|_2^4 \|\boldsymbol{y}_j - \boldsymbol{x}_0\|_2^4,$$

$$\mathcal{R}_3(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m) = \sum_{i=1}^m \|\boldsymbol{y}_i - \boldsymbol{x}_0\|_2^4$$

only depend on the given interpolation points $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m$, given a base point $\boldsymbol{x}_0$ at the current iteration, and $\mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3$ are functions of $C_1, C_2, C_1^*, C_2^*, n, r$.

Theorem. If $\mathcal{R}_1(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_m) \to 0$, then $C_1 = \frac{1-\varepsilon}{3}, C_2 = \frac{1-\varepsilon}{3}$ is the couple of weight coefficients providing the central KKT matrix, where ε is the lower bound of C_3.

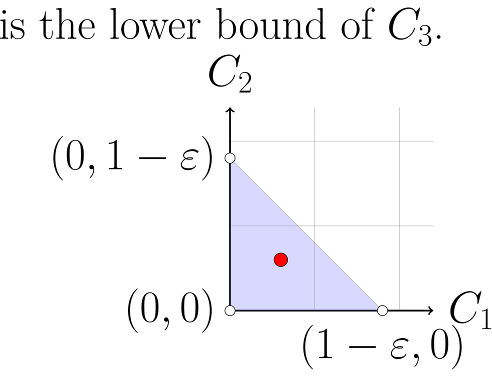


Figure 2:Coefficient region $\mathcal{C}$

Numerical Results

Example of the barycenter

Table 1:$\text{Error}_{\text{ave}}^{(1)}$ and $\text{Error}_{\text{ave}}^{(2)}$ for sampled coefficients, $\varepsilon = 0.01$, $n = 100$

(C_1, C_2)	$(\frac{1-\varepsilon}{3}, \frac{1-\varepsilon}{3})$	$(\frac{1}{2}-\varepsilon, \frac{1}{2})$	$(0, \frac{1}{2})$	$(1-\varepsilon, 0)$	$(0, 1-\varepsilon)$	$(0,0)$
$\text{Error}_{\text{ave}}^{(1)}$	5.663	8.655	8.988	18.3	49.66	16.99
$\text{Error}_{\text{ave}}^{(2)}$	2.467	136.2	2.611	136.2	136.2	2.795

Table 1 numerically supports that $(\frac{1-\varepsilon}{3}, \frac{1-\varepsilon}{3})$ has the `smallest` $\text{Error}_{\text{ave}}$ among the six pairs of weight coefficients.

Comparison for solving an example

We apply derivative-free algorithms based on a least weighted H^2 norm updating quadratic model with different weight coefficients to minimize the `2-dimensional Rosenbrock` function

$$f(\boldsymbol{y}) = (1-y_1)^2 + 100(y_2 - y_1^2)^2, \tag{4}$$

which is a smooth nonconvex function. The minimizer of (4) is $\boldsymbol{y}_{\min} = (1.04, 1.1)^\top$ and lies on a narrow, curved valley.

Table 2:Different scaled (semi-)norms with corresponding coefficients

Weight coeffients	Corresponding scaled (semi-)norms	Norm No.
$C_1 = \frac{1}{3}$ $C_2 = \frac{1}{3}$ $C_3 = \frac{1}{3}$	H^2 norm	(a)
$C_1 = \frac{1}{2}$ $C_2 = \frac{1}{2}$ $C_3 = 0$	H^1 norm	(b)
$C_1 = 0$ $C_2 = \frac{1}{2}$ $C_3 = \frac{1}{2}$	H^1 semi-norm + H^2 semi-norm	(c)
$C_1 = 1$ $C_2 = 0$ $C_3 = 0$	H^0 norm (L^2 norm)	(d)
$C_1 = 0$ $C_2 = 1$ $C_3 = 0$	H^1 semi-norm	(e)
$C_1 = 0$ $C_2 = 0$ $C_3 = 1$	H^2 semi-norm (Frobenius norm of Hessian)	(f)

Table 3:Rosenbrock approximate solutions found after 16 evaluations

Norm No.	Function value	Solution
(a)	0.0031	$(1.0495, 1.1040)^\top$
(b)	0.0169	$(1.0427, 1.0996)^\top$
(c)	0.0203	$(1.0418, 1.0990)^\top$
(d)	0.0078	$(1.0455, 1.1008)^\top$
(e)	0.0169	$(1.0427, 1.0996)^\top$
(f)	0.0147	$(1.0426, 1.0984)^\top$

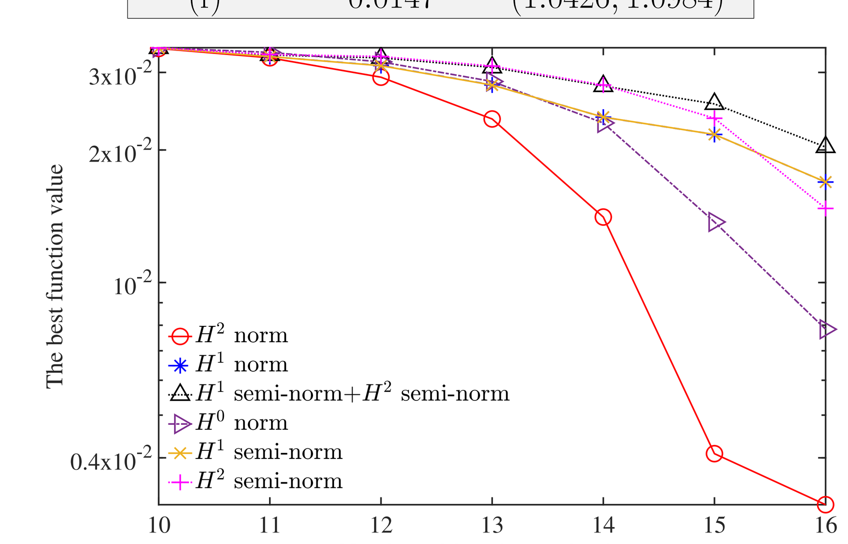


Figure 3:Minimizing the Rosenbrock function with different models

Performance profiles on a larger test set

Table 4:60 nonlinear, nonconvex test problems

ARGLINA	ARGLINA4	ARGLINB	ARGLINC	ARGTRIG	ARWHEAD
BDQRTIC	BDQRTICP	BDVALUE	BROWNAL	BROYDN3D	BROYDN7D
BRYBND	CHAINWOO	CHNROSNB	CHPOWELLB	CHROSEN	CRAGGLVY
CUBE	DQRTIC	EDENSCH	ENGVAL1	ERRINROS	EXPSUM
EXTROSNB	EXTTET	FIROSE	FLETCBV2	FLETCBV3	FLETCHCR
FREUROTH	GENBROWN	GENROSE	INDEF	INTEGREQ	LIARWHD
LILIFUN3	LILIFUN4	MOREBV	MOREBVL	NONDIA	PENALTY1
PENALTY2	PENALTY3	PENALTY3P	ROSENBROCK	SBRYBND	SBRYBNDL
SEROSE	SINQUAD	SROSENBR	STMOD	TOINTTRIG	TQUARTIC
TRIGSABS	TRIGSSQS	TRIROSE1	TRIROSE2	VARDIM	WOODS

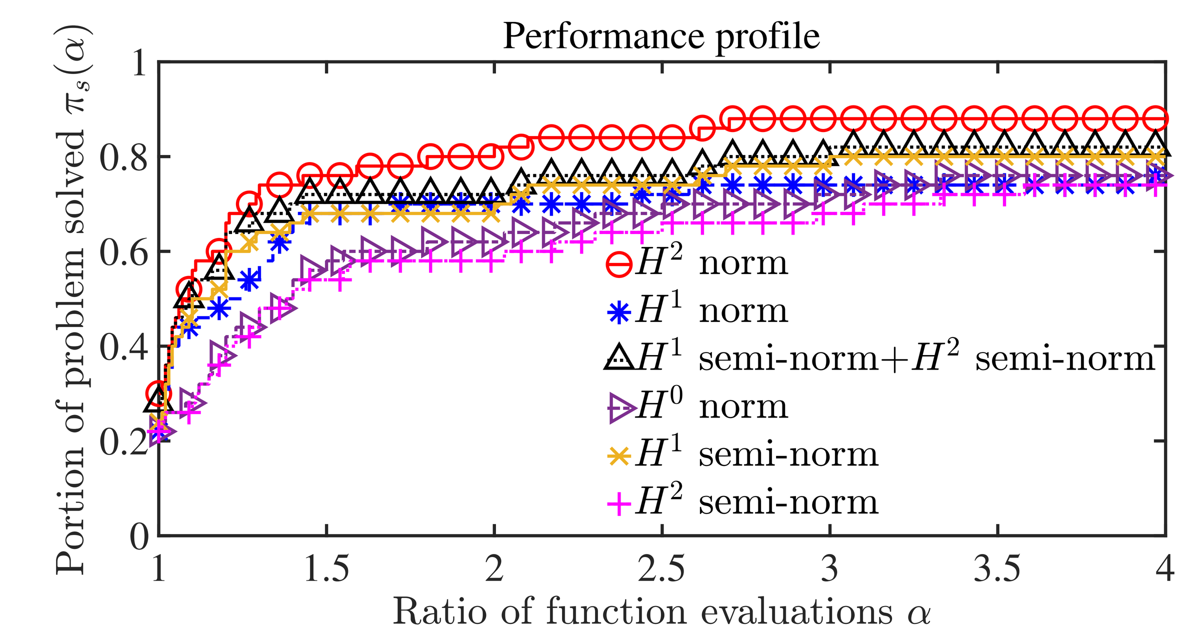


Figure 4:Solving test problems in Table 4 with different models. Higher and to the left indicates better performance in a performance profile.

Conclusion and Next Steps

This poster proposes ways to define quadratic models using different weight coefficients in a weighted H^2 norm updating interpolation quadratic models. We evaluate how the choice of coefficients affects the performance of derivative-free trust-region algorithms. We find that the barycenter of the weight coefficient region $\mathcal{C}$ performs especially well. In ongoing work, we evaluate additional properties of these models and their use in optimization algorithms, with a goal of delivering robust performance across a wide range of computational budgets.

References

[1] J. Larson, M. Menickelly, and S. M. Wild, *Derivative-free optimization methods*, Acta Numer., 28 (2019), pp. 287–404.

[2] M. J. D. Powell, *On trust region methods for unconstrained minimization without derivatives*, Math. Program., 97 (2003), pp. 605–623.

[3] S. M. Wild, *MNH: A derivative-free optimization algorithm using minimal norm Hessians*, in The Tenth Copper Mountain Conference on Iterative Methods, 2008.

[4] S. M. Wild and C. Shoemaker, *Global convergence of radial basis function trust region derivative-free algorithms*, SIAM J. Optim., 21 (2011), pp. 761–781.

[5] P. Xie and Y.-x. Yuan, *Least H^2 norm updating quadratic interpolation model function for derivative-free trust-region algorithms*, to appear in IMA J. Numer. Anal., (2024).

Acknowledgments

This work was supported by LBNL LDRD and the DOE Office of Science, Office of Advanced Scientific Computing Research applied math program under Contract DE-AC02-05CH11231. The first author also thanks SIAM for a travel support award.