COBYQA

A derivative-free trust-region SQP method for nonlinearly constrained optimization

Tom M. Ragonneau Zaikun Zhang

CSE23 (March 2, 2023)

Department of Applied Mathematics
The Hong Kong Polytechnic University

We design a method named COBYQA for solving[1]

$$
\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & f(x) \\
\text{s.t.} \quad & c(x) \leq 0, \\
& l \leq x \leq u,
\end{aligned}
$$

where derivatives of f and c are unavailable.

Notes on the method

- COBYQA aims at being a successor to COBYLA (Powell 1994).
- We implement COBYQA into a Python solver.
- The bound constraints are handled separately.

[1] The equality constraints are omitted here for simplicity.

Inviolable bound constraints

The bound constraints $l \leq x \leq u$ are assumed inviolable

- They often represent inalienable restrictions.
- f or c may not be defined otherwise.

Therefore, COBYQA always respects the bound constraints.

A few examples from academia and industry

- Optimization method tuning (Audet and Orban 2006).
- Hyperparameter tuning (Ghanbari and Scheinberg 2017).
- Aircraft engineering (Gazaix et al. 2019).

Table of contents

The general framework

The derivative-free trust-region SQP method

COBYQA iteratively solves the trust-region SQP subproblem

$$
\begin{aligned}
\min_{d \in \mathbb{R}^n} \quad & \nabla f(x^k)^\mathsf{T} d + \frac{1}{2} d^\mathsf{T} \nabla^2_{x,x} \mathcal{L}(x^k, \lambda^k) d \\
\text{s.t.} \quad & c(x^k) + \nabla c(x^k) d \le 0, \\
& l \le x^k + d \le u, \\
& \|d\| \le \Delta^k,
\end{aligned}
$$

with $\mathcal{L}(x, \lambda) = f(x) + \lambda^\mathsf{T} c(x)$.

The derivative-free trust-region SQP method

COBYQA iteratively solves the trust-region SQP subproblem

$$\begin{aligned}
\min_{d \in \mathbb{R}^n} \quad & \nabla \hat{f}^k(x^k)^\mathsf{T} d + \frac{1}{2} d^\mathsf{T} \nabla^2_{x,x} \hat{\mathcal{L}}^k(x^k, \lambda^k) d \\
\text{s.t.} \quad & \hat{c}^k(x^k) + \nabla \hat{c}^k(x^k) d \le 0, \\
& l \le x^k + d \le u, \\
& \|d\| \le \Delta^k,
\end{aligned}$$

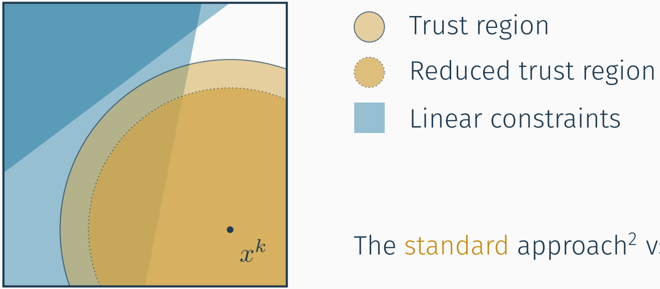
with $\hat{\mathcal{L}}^k(x, \lambda) = \hat{f}^k(x) + \lambda^\mathsf{T} \hat{c}^k(x)$, given some models $\hat{f}^k$ and $\hat{c}^k$.

Remarks on this subproblem

- We only require an approximate solution d^k.
- The solution must satisfy $l \le x^k + d^k \le u$.
- The subproblem may be infeasible. What is a solution?

We compute $d^k = n^k + t^k$, where

- the normal step n^k reduces the (possible) constraint violation, and
- the tangential step t^k reduces the quadratic objective function.

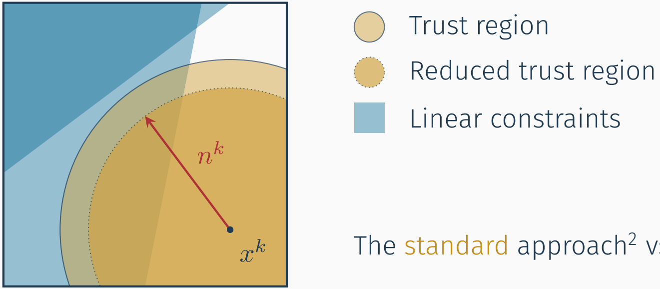


The standard approach[2] vs. the new one.

[2] Conn, Gould, and Toint (2000, §15.4.4).

We compute $d^k = n^k + t^k$, where

- the normal step n^k reduces the (possible) constraint violation, and
- the tangential step t^k reduces the quadratic objective function.

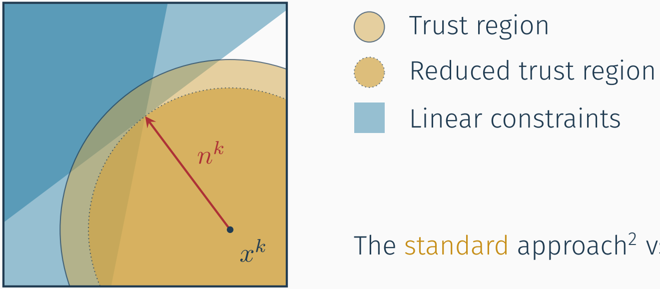


The standard approach[2] vs. the new one.

[2] Conn, Gould, and Toint (2000, §15.4.4).

We compute $d^k = n^k + t^k$, where

- the normal step n^k reduces the (possible) constraint violation, and
- the tangential step t^k reduces the quadratic objective function.



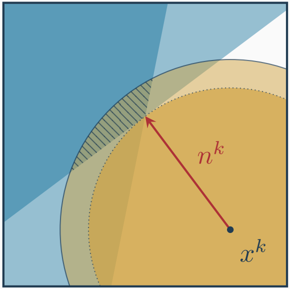
The standard approach[2] vs. the new one.

[2] Conn, Gould, and Toint (2000, §15.4.4).

A new Byrd-Omojokun approach

We compute $d^k = n^k + t^k$, where

- the normal step n^k reduces the (possible) constraint violation, and
- the tangential step t^k reduces the quadratic objective function.



Trust region

Reduced trust region

Linear constraints

Feasible region for t^k

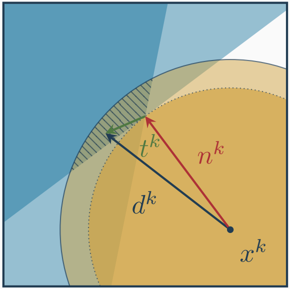
The standard approach[2] vs. the new one.

[2] Conn, Gould, and Toint (2000, §15.4.4).

A new Byrd-Omojokun approach

We compute $d^k = n^k + t^k$, where

- the normal step n^k reduces the (possible) constraint violation, and
- the tangential step t^k reduces the quadratic objective function.



- Trust region
- Reduced trust region
- Linear constraints
- Feasible region for t^k

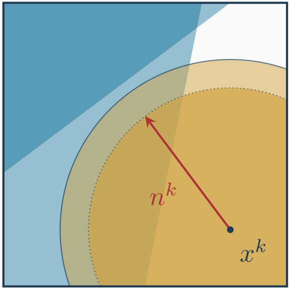
The standard approach[2] vs. the new one.

[2] Conn, Gould, and Toint (2000, §15.4.4).

A new Byrd-Omojokun approach

We compute $d^k = n^k + t^k$, where

- the normal step n^k reduces the (possible) constraint violation, and
- the tangential step t^k reduces the quadratic objective function.



- Trust region
- Reduced trust region
- Linear constraints
- Feasible region for t^k

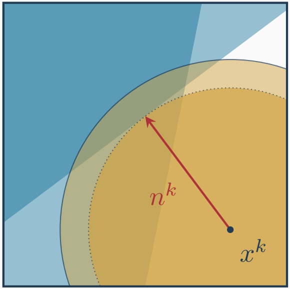
The standard approach[2] vs. the new one.

[2] Conn, Gould, and Toint (2000, §15.4.4).

A new Byrd-Omojokun approach

We compute $d^k = n^k + t^k$, where

- the normal step n^k reduces the (possible) constraint violation, and
- the tangential step t^k reduces the quadratic objective function.



Trust region

Reduced trust region

Linear constraints

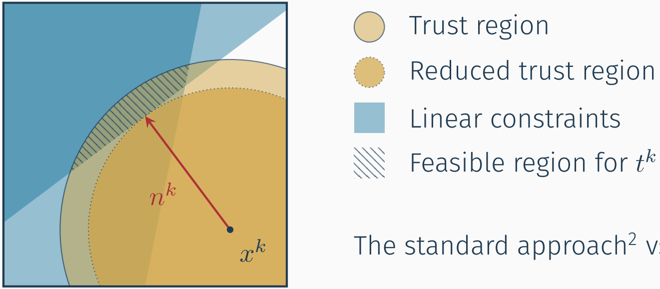
Feasible region for t^k

The standard approach[2] vs. the new one.

[2] Conn, Gould, and Toint (2000, §15.4.4).

We compute $d^k = n^k + t^k$, where

- the normal step n^k reduces the (possible) constraint violation, and
- the tangential step t^k reduces the quadratic objective function.

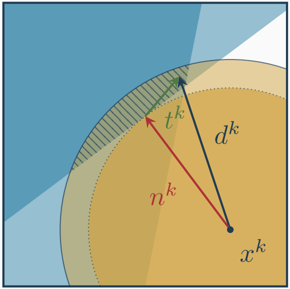


The standard approach[2] vs. the new one.

[2] Conn, Gould, and Toint (2000, §15.4.4).

We compute $d^k = n^k + t^k$, where

- the normal step n^k reduces the (possible) constraint violation, and
- the tangential step t^k reduces the quadratic objective function.



Trust region

Reduced trust region

Linear constraints

Feasible region for t^k

The standard approach[2] vs. the new one.

The feasible region for t^k is wider in the new approach.

[2] Conn, Gould, and Toint (2000, §15.4.4).

The standard approach is as follows.

- The normal step n^k solves approximately (for some $\zeta < 1$)

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \left\| \left[\hat{c}^k(x^k) + \nabla \hat{c}^k(x^k) d \right]^+ \right\| \\ \text{s.t.} \quad & l \le x^k + d \le u, \\ & \|d\| \le \zeta \Delta^k. \end{aligned}$$

- The tangential step t^k solves approximately

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \left[\nabla \hat{f}^k(x^k) + \nabla^2_{x,x} \hat{\mathcal{L}}^k(x^k, \lambda^k) n^k \right]^{\mathsf{T}} d + \frac{1}{2} d^{\mathsf{T}} \nabla^2_{x,x} \hat{\mathcal{L}}^k(x^k, \lambda^k) d \\ \text{s.t.} \quad & \nabla \hat{c}^k(x^k)^{\mathsf{T}} d \le 0, \\ & l \le x^k + n^k + d \le u, \\ & \|n^k + d\| \le \Delta^k. \end{aligned}$$

A new Byrd-Omojokun approach (cont'd)

The new approach is as follows.

- The normal step n^k solves approximately (for some $\zeta < 1$)

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \left\| \left[\hat{c}^k(x^k) + \nabla \hat{c}^k(x^k) d \right]^+ \right\| \\ \text{s.t.} \quad & l \leq x^k + d \leq u, \\ & \|d\| \leq \zeta \Delta^k. \end{aligned}$$

- The tangential step t^k solves approximately

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \quad & \left[\nabla \hat{f}^k(x^k) + \nabla^2_{x,x} \hat{\mathcal{L}}^k(x^k, \lambda^k) n^k \right]^\mathsf{T} d + \frac{1}{2} d^\mathsf{T} \nabla^2_{x,x} \hat{\mathcal{L}}^k(x^k, \lambda^k) d \\ \text{s.t.} \quad & \nabla \hat{c}^k(x^k)^\mathsf{T} d \leq [\hat{c}^k(x^k) + \nabla \hat{c}^k(x^k) n^k]^-, \\ & l \leq x^k + n^k + d \leq u, \\ & \|n^k + d\| \leq \Delta^k. \end{aligned}$$

Interpolation-based models

COBYQA models f and c by quadratic interpolation, as follows.[3]

Derivative-free symmetric Broyden update (Powell 2004)

The kth model $\hat{f}^k$ of f solves

$$\begin{aligned} \min_{Q} \quad & \left\| \nabla^2 \hat{f}^{k-1} - \nabla^2 Q \right\|_{\mathsf{F}} \\ \text{s.t.} \quad & Q(y) = f(y), \ y \in \mathcal{Y}^k, \end{aligned}$$

for some $\mathcal{Y}^k \subseteq \mathbb{R}^n$, with $\hat{f}^{-1} \equiv 0$. The model $\hat{c}^k$ of c is built similarly.

The interpolation set $\mathcal{Y}^k$ is recycled at each iteration.

- The set $\mathcal{Y}^{k+1}$ is built as $(\mathcal{Y}^k \setminus \bar{y}) \cup \{x^k + d^k\}$ for some $\bar{y} \in \mathcal{Y}^k$.
- The KKT system of this variational problem is linear.

[3] Other methods: Conn, Scheinberg, and Toint (1997a,b, 1998), Wild (2008), Bandeira, Scheinberg, and Vicente (2012), Zhang (2014), Xie and Yuan (2023).

Many difficulties arise

A lot of questions must be addressed

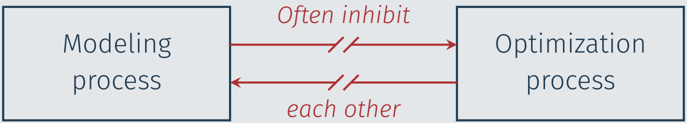
- How to calculate the steps n^k and t^k numerically?
  COBYQA adapts the truncated conjugate gradient method.
- What is the approximate Lagrange multiplier λ^k?
  We choose the least-squares Lagrange multiplier.
- Which merit function should we use?
  COBYQA uses the ℓ_2-merit function.
- How to update the penalty parameter?
  The update incorporates
  - a theoretical value for the exactness of the merit function, and
  - a strategy used by Powell in COBYLA.

These questions (and more) are addressed in Ragonneau (2022).

A crucial difficulty in the implementation

- What if the interpolation set $\mathcal{Y}^k$ is almost nonpoised?
  A well-known approach: using a geometry-improving mechanism.[4]

This is a central difficulty in the implementation of DFO methods



- The iterates $\{x^k\}$ likely lie on a particular path.
- The modeling process does not ponder the optimization problem.

[4] Conn, Scheinberg, and Vicente (2008a,b), Fasano, Morales, and Nocedal (2009).

Implementation and experiments

The Python implementation of COBYQA

A quote from Powell (2006)

> "The development of NEWUOA has taken nearly three years. The work was very frustrating [...]"

The development of COBYQA was not easier.

We implemented COBYQA in Python and made it publicly available.

Documentation

Source code

Comparing COBYQA with existing DFO solvers

- We assess the quality of points based on

$$\varphi(x) = \begin{cases} f(x) & \text{if } v_\infty(x) \le 10^{-10}, \\ \infty & \text{if } v_\infty(x) \ge 10^{-5}, \\ f(x) + 10^5 v_\infty(x) & \text{otherwise,} \end{cases}$$

  where v_∞ denotes the ℓ_∞-constraint violation.
- The problems are from the CUTEst set.
- The problems are of dimension at most 50 (this is not small).
- The noisy problems replace f with

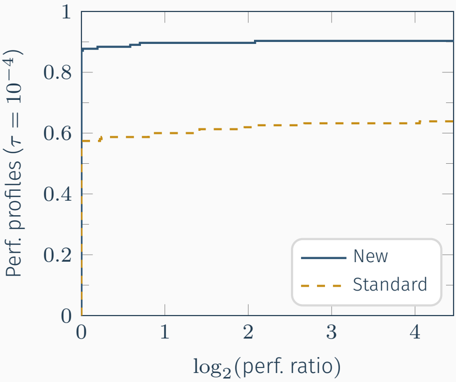
$$\tilde{f}(x) = \big[1 + \epsilon(x)\big] f(x),$$

  where $\epsilon(x) \sim \mathcal{N}(0, \sigma^2)$.

Performance of the new Byrd-Omojokun approach

We compare the new and the standard Byrd-Omojokun approaches

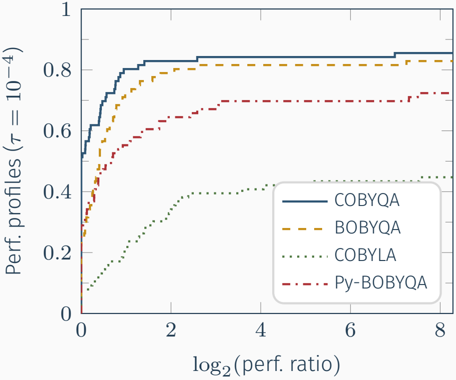
- on linearly and nonlinearly constrained problems,
- in the implementation of COBYQA.



Performance on bound-constrained problems

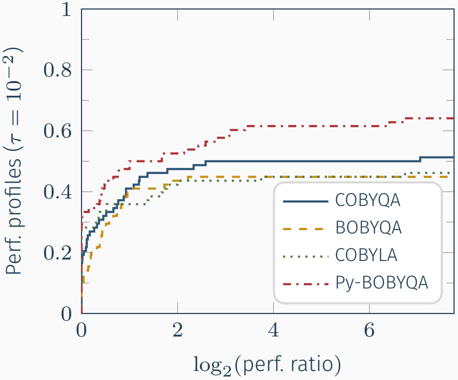
We compare COBYQA, COBYLA, and two implementations of BOBYQA

- on bound-constrained problems,



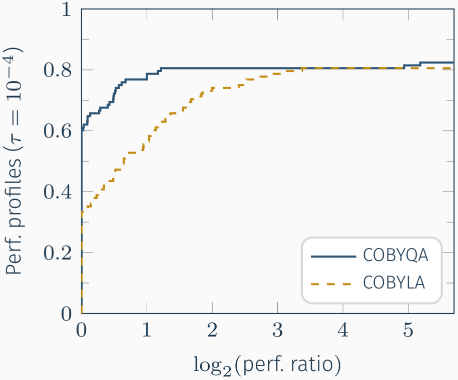
We compare COBYQA, COBYLA, and two implementations of BOBYQA

- on bound-constrained problems,
- adding noise to f, with $\sigma = 10^{-3}$.



We compare COBYQA and COBYLA

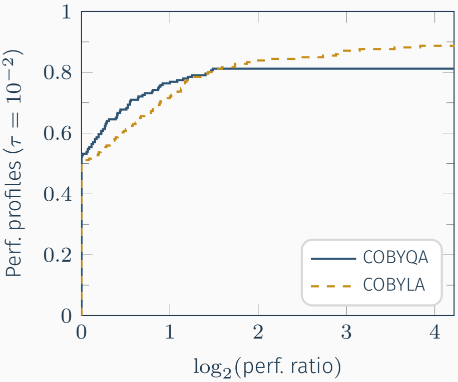
- on nonlinearly constrained problems,



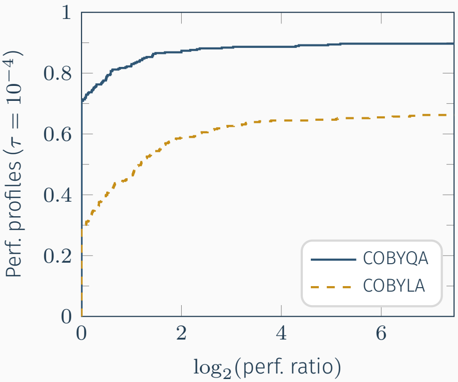
Performance on nonlinearly constrained problems

We compare COBYQA and COBYLA

- on nonlinearly constrained problems,
- adding noise to f, with $\sigma = 10^{-3}$.



We compare COBYQA and COBYLA on all problems.



Conclusion

Conclusion

- COBYQA already received positive feedback from practitioners.
- It will soon be included in
  - PDFO as a successor for COBYLA, and
  - GEMSEO, an industrial software package for MDO.
- We will soon investigate the convergence properties of COBYQA.

For more information, visit:

COBYQA's website

My website

My thesis

References I

- Audet, C. and Orban, D. (2006). "Finding optimal algorithmic parameters using derivative-free optimization". *SIAM J. Optim.* 17, pp. 642–664.
- Bandeira, A. S., Scheinberg, K., and Vicente, L. N. (2012). "Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization". *Math. Program.* 134, pp. 223–257.
- Byrd, R. H. (1987). "Robust trust region methods for constrained optimization". In: *The Third SIAM Conference on Optimization*.
- Conn, A. R., Gould, N. I. M., and Toint, Ph. L. (2000). *Trust-Region Methods*. MPS-SIAM Ser. Optim. Philadelphia, PA, US: SIAM.

References II

- Conn, A. R., Scheinberg, K., and Toint, Ph. L. (1997a). "On the convergence of derivative-free methods for unconstrained optimization". In: *Approximation Theory and Optimization: Tributes to M. J. D. Powell*. Ed. by M. D. Buhmann and A. Iserles. Cambridge, UK: Cambridge University Press, pp. 83–108.
- — (1997b). "Recent progress in unconstrained nonlinear optimization without derivatives". *Math. Program.* 79, pp. 397–414.
- — (1998). "A derivative free optimization algorithm in practice". In: *Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*. St. Louis, MO, US: AIAA, pp. 129–139.

References III

- Conn, A. R., Scheinberg, K., and Vicente, L. N. (2008a). “Geometry of interpolation sets in derivative free optimization”. *Math. Program.* 111, pp. 141–172.
- — (2008b). “Geometry of sample sets in derivative-free optimization: polynomial regression and underdetermined interpolation”. *IMA J. Numer. Anal.* 28, pp. 721–748.
- Fasano, G., Morales, J. L., and Nocedal, J. (2009). “On the geometry phase in model-based algorithms for derivative-free optimization”. *Optim. Methods Softw.* 24, pp. 145–154.
- Gazaix, A. et al. (2019). “Industrial application of an advanced bi-level MDO formulation to aircraft engine pylon optimization”. In: *AIAA Aviation Forum*. Dallas, TX, US: AIAA.

References iv

- Ghanbari, H. and Scheinberg, K. (2017). *Black-box optimization in machine learning with trust region based derivative free algorithm*. Tech. rep. 17T-005. Bethlehem, PA, US: COR@L.
- Omojokun, E. O. (1989). “Trust Region Algorithms for Optimization with Nonlinear Equality and Inequality Constraints”. Ph.D. thesis. Boulder, CO, US: University of Colorado Boulder.
- Powell, M. J. D. (1994). “A direct search optimization method that models the objective and constraint functions by linear interpolation”. In: *Advances in Optimization and Numerical Analysis*. Ed. by S. Gomez and J. P. Hennart. Dordrecht, NL: Springer, pp. 51–67.
- — (2004). “Least Frobenius norm updating of quadratic models that satisfy interpolation conditions”. *Math. Program.* 100, pp. 183–215.

- Powell, M. J. D. (2006). "The NEWUOA software for unconstrained optimization without derivatives". In: *Large-Scale Nonlinear Optimization*. Ed. by G. Di Pillo and M. Roma. New York, NY, US: Springer, pp. 255–297.
- Ragonneau, T. M. (2022). "Model-Based Derivative-Free Optimization Methods and Software". Ph.D. thesis. Hong Kong: Department of Applied Mathematics, The Hong Kong Polytechnic University.
- Wild, S. M. (2008). "MNH: a derivative-free optimization algorithm using minimal norm Hessians". In: *The Tenth Copper Mountain Conference on Iterative Methods*.
- Xie, P. and Yuan, Y. (2023). *Least H^2 norm updating quadratic interpolation model function for derivative-free trust-region algorithms*. arXiv: `2302.12017 [math.OC]`.

- Zhang, Z. (2014). “Sobolev seminorm of quadratic functions with applications to derivative-free optimization”. *Math. Program.* 146, pp. 77–96.

The source code of this presentation is available at

github.com/ragonneau/cse23.

It is based on the metropolis Beamer theme, available at

github.com/matze/mtheme.