COBYQA

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

Tom M. Ragonneau Zaikun Zhang

SIAM Conference on Optimization, Seattle, WA, 2023

Department of Applied Mathematics
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

This work was supported by the Hong Kong PhD Fellowship Scheme.

We design a method named COBYQA for solving

$$\min_{x \in \mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad \begin{cases} g(x) \leq 0, \\ h(x) = 0, \\ l \leq x \leq u, \end{cases}$$

where derivatives of f, g, and h 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 assumed inviolable.
  - They often represent inalienable restrictions.
  - The functions f, g, and h may not be defined outside the bounds.

Table of contents

General framework of COBYQA

COBYQA iteratively solves the trust-region SQP subproblem

$$\min_{s \in \mathbb{R}^n} \quad f(x_k) + \nabla f(x_k)^\mathsf{T} s + \frac{1}{2} s^\mathsf{T} \nabla^2_{x,x} \mathcal{L}(x_k, \lambda_k, \mu_k) s$$

$$\text{s.t.} \quad \begin{cases} g(x_k) + \nabla g(x_k) s \le 0, \\ h(x_k) + \nabla h(x_k) s = 0, \\ l \le x_k + s \le u, \\ \|s\| \le \Delta_k, \end{cases}$$

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

The derivative-free trust-region SQP method

COBYQA iteratively solves the trust-region SQP subproblem

$$\min_{s \in \mathbb{R}^n} \quad \hat{f}_k(x_k) + \nabla \hat{f}_k(x_k)^{\mathsf{T}} s + \frac{1}{2} s^{\mathsf{T}} \nabla^2_{x,x} \widehat{\mathcal{L}}_k(x_k, \lambda_k, \mu_k) s$$

$$\text{s.t.} \quad \begin{cases} \hat{g}_k(x_k) + \nabla \hat{g}_k(x_k) s \le 0, \\ \hat{h}_k(x_k) + \nabla \hat{h}_k(x_k) s = 0, \\ l \le x_k + s \le u, \\ \|s\| \le \Delta_k, \end{cases}$$

with $\widehat{\mathcal{L}}_k(x, \lambda, \mu) = \hat{f}_k(x) + \lambda^{\mathsf{T}} \hat{g}_k(x) + \mu^{\mathsf{T}} \hat{h}_k(x)$, given some models.

Remarks on this subproblem

- We only require an approximate solution s_k.
- The solution must satisfy $l \le x_k + s_k \le u$.
- It is wrong to replace $\nabla^2_{x,x} \widehat{\mathcal{L}}_k(x_k, \lambda_k, \mu_k)$ with $\nabla^2 \hat{f}_k(x_k)$.

COBYQA models f, g, and h by quadratic interpolation, as follows.[1]

Derivative-free symmetric Broyden update (Powell 2004)

The kth quadratic model $\hat{f}_k$ of f solves

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

for some $\mathcal{Y}_k \subseteq \mathbb{R}^n$, with $\hat{f}_{-1} \equiv 0$ (similar for $\hat{g}_k$ and $\hat{h}_k$).

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

- We set $\mathcal{Y}_{k+1} = \left(\mathcal{Y}_k \cup \{x_k + s_k\}\right) \setminus \{\bar{y}\}$ for some bad point $\bar{y} \in \mathcal{Y}_k$.
- This variational problem is a QP, with a linear KKT system.

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

Management of the trust-region radius

We maintain Δ_k and a lower bound $\delta_k \le \Delta_k$

- The lower bound δ_k is never increased.
- We update Δ_k in the usual way, but we always have $\Delta_k \ge \delta_k$.
- This strategy is adapted from Powell's methods (e.g., NEWUOA).

The value of δ_k is an indicator of the current resolution.

- Without $\Delta_k \ge \delta_k$, the value of Δ_k may become too small.
- It prevents the interpolation points from concentrating too much.
- The value of δ_k is only decreased when necessary.
- Hence, stopping when $\delta_k \le \delta_{\text{end}}$ is reasonable ($\delta_{\text{end}} > 0$).

For more information, see Ragonneau (2022, § 5.2.5).

There remain many difficulties to address

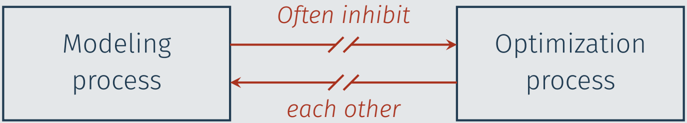
- What if the trust-region subproblem is infeasible?
  COBYQA uses a Byrd-Omojokun composite-step approach.
- How to calculate the trial step s_k numerically?
  We adapt the truncated conjugate gradient method.
- What are the approximate Lagrange multipliers λ_k and μ_k?
  We choose the least-squares Lagrange multipliers.
- How to define a trust-region ratio? Using what merit function?
  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.[2]

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.

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

A new interpretation of SQP

⚠ This part is NOT about DFO but SQP.

For simplicity, we consider the smooth equality-constrained problem

$$\min_{x \in \mathbb{R}^n} \quad f(x) \quad \text{s.t.} \quad h(x) = 0,$$

and we denote by $\mathcal{L}$ its Lagrangian, given by

$$\mathcal{L}(x, \mu) = f(x) + \mu^\mathsf{T} h(x).$$

Classical interpretations of SQP

- It is Newton's method applied to the KKT system.
- It iteratively minimizes the second-order Taylor expansion of

$$\widetilde{\mathcal{L}}(x, \mu) = f(x) + \mu^\mathsf{T} \big[h(x) - h(x_k) - \nabla h(x_k)(x - x_k) \big].$$

- It is equivalent to a convex-composite algorithm for NLP.

The SQP subproblem at $(\bar{x}, \bar{\mu})$ is

$$\begin{aligned} \min_{s \in \mathbb{R}^n} \quad & f(\bar{x}) + \nabla f(\bar{x})^\mathsf{T} s + \frac{1}{2} s^\mathsf{T} \nabla^2_{x,x} \mathcal{L}(\bar{x}, \bar{\mu}) s \\ \text{s.t.} \quad & h(\bar{x}) + \nabla h(\bar{x}) s = 0. \end{aligned}$$

A curve on the level surface of the constraints

We consider a curve parametrized by $\xi : \mathbb{R} \to \mathbb{R}^n$ with

$$h\big(\xi(t)\big) = h(\bar{x}) \quad \text{for all } t \in \mathbb{R} \quad \text{and} \quad \xi(0) = \bar{x}.$$

Note that $\bar{x}$ can be infeasible, i.e., $h(\bar{x}) \neq 0$.

Remark that $\nabla h(\bar{x}) \xi'(0) = 0$, i.e., $\xi'(0) \in \ker \nabla h(\bar{x})$.

The objective function of the SQP subproblem at $(\bar{x}, \bar{\mu})$ is

$$q(s) = f(\bar{x}) + \nabla f(\bar{x})^\mathsf{T} s + \frac{1}{2} s^\mathsf{T} \nabla^2_{x,x} \mathcal{L}(\bar{x}, \bar{\mu}) s.$$

Main result (Ragonneau and Zhang, 2022)

If f, h, and ξ have locally Lipschitz second-order derivatives, then

$$\big| f\big(\xi(t)\big) - q\big(\xi'(0)t\big) \big| \le \left(\nu t + \frac{1}{2} \big| \xi''(0)^\mathsf{T} \nabla_x \mathcal{L}(\bar{x}, \bar{\mu}) \big| \right) t^2$$

for some $\nu \ge 0$, $\epsilon > 0$, and all $t \in (-\epsilon, \epsilon)$.

The new interpretation (cont'd)

The objective function of the SQP subproblem at $(\bar{x}, \bar{\mu})$ is

$$q(s) = f(\bar{x}) + \nabla f(\bar{x})^{\mathsf{T}} s + \frac{1}{2} s^{\mathsf{T}} \nabla^2 f(\bar{x}) s.$$

Main result (Ragonneau and Zhang, 2022)

If f, h, and ξ have locally Lipschitz second-order derivatives, then

$$\big| f\big(\xi(t)\big) - q\big(\xi'(0)t\big) \big| \le \left(\nu t + \frac{1}{2} \big| \xi''(0)^{\mathsf{T}} \nabla f(\bar{x}) \big| \right) t^2$$

for some $\nu \ge 0$, $\epsilon > 0$, and all $t \in (-\epsilon, \epsilon)$.

The new interpretation (cont'd)

The objective function of the SQP subproblem at $(\bar{x}, \bar{\mu})$ is

$$q(s) = f(\bar{x}) + \nabla f(\bar{x})^{\mathsf{T}} s + \frac{1}{2} s^{\mathsf{T}} \nabla^2_{x,x} \mathcal{L}(\bar{x}, \bar{\mu}) s.$$

Main result (Ragonneau and Zhang, 2022)

If f, h, and ξ have locally Lipschitz second-order derivatives, then

$$\left| f\big(\xi(t)\big) - q\big(\xi'(0)t\big) \right| \le \left(\nu t + \frac{1}{2} \left| \xi''(0)^{\mathsf{T}} \nabla_x \mathcal{L}(\bar{x}, \bar{\mu}) \right| \right) t^2$$

for some $\nu \ge 0$, $\epsilon > 0$, and all $t \in (-\epsilon, \epsilon)$.

- Note that $\nabla_x \mathcal{L}(\bar{x}, \bar{\mu}) \approx 0$ if $(\bar{x}, \bar{\mu})$ is almost a KKT pair.

The objective function of the SQP subproblem at $(\bar{x}, \bar{\mu})$ is

$$q(s) = f(\bar{x}) + \nabla f(\bar{x})^\mathsf{T} s + \frac{1}{2} s^\mathsf{T} \nabla^2_{x,x} \mathcal{L}(\bar{x}, \bar{\mu}) s.$$

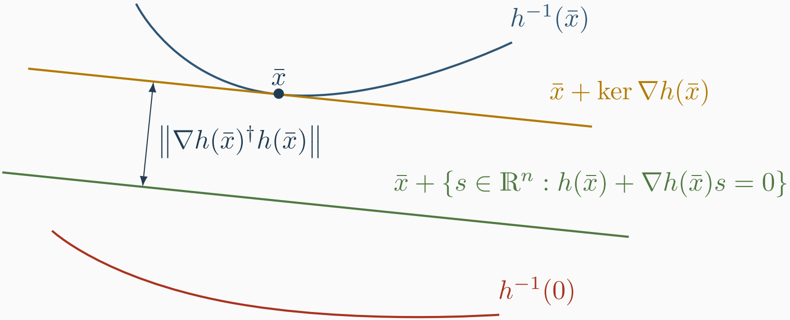
Main result (Ragonneau and Zhang, 2022)

If f, h, and ξ have locally Lipschitz second-order derivatives, then

$$\left| f\big(\xi(t)\big) - q\big(\xi'(0)t\big) \right| \le \left(\nu t + \frac{1}{2} \left| \xi''(0)^\mathsf{T} \nabla_x \mathcal{L}(\bar{x}, \bar{\mu}) \right| \right) t^2$$

for some $\nu \ge 0$, $\epsilon > 0$, and all $t \in (-\epsilon, \epsilon)$.

- Note that $\nabla_x \mathcal{L}(\bar{x}, \bar{\mu}) \approx 0$ if $(\bar{x}, \bar{\mu})$ is almost a KKT pair.
- What does this theorem mean geometrically?



- The green line represents the feasible set of the SQP subproblem.
- If $h(\bar{x}) = 0$, the green and yellow lines overlap.
- The green line is a shifted copy of the yellow one towards feasibility.

Implementation and experiments

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.

`www.cobyqa.com`

Installation via PyPI

```
$ pip install cobyqa
```

Solving SOLAR6 (Garneau 2015)

The problem has

- 5 continuous variables,
- inviolable bound constraints on each variable,
- 6 nonlinear inequality constraints, and
- an infeasible initial guess, with $\|\max\{0, g(x_0)\}\|_\infty \approx 44.9965$.

Output of COBYQA (with default options)

- Number of function evaluations: 192.
- Final objective function value: $f(x_{192}) = 54\,842\,275.4721$.
- Final constraint violation: $\|\max\{0, g(x_{192})\}\|_\infty = 0.0$.

The best objective function value so far is $f(x_*) = 43\,955\,452.8547$.

Comparing COBYQA with existing DFO solvers

- We assess the quality of points based on the merit function

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

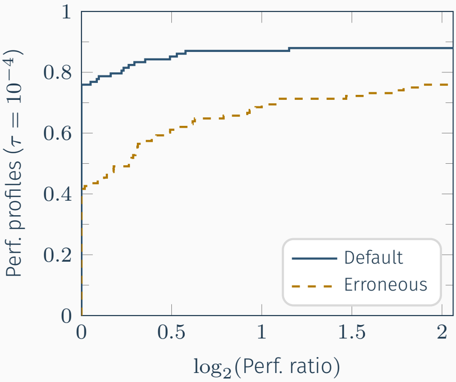
  where v_{∞} denotes the ℓ_{∞}-constraint violation.
- The problems are from the CUTEst set.
- The problems are of dimension at most 50 (this is not small).
- Problems with inviolable bounds replace f with

$$\tilde{f}(x) = \begin{cases} f(x) & \text{if } l \le x \le u, \\ \infty & \text{otherwise.} \end{cases}$$

Performance of the SQP approach

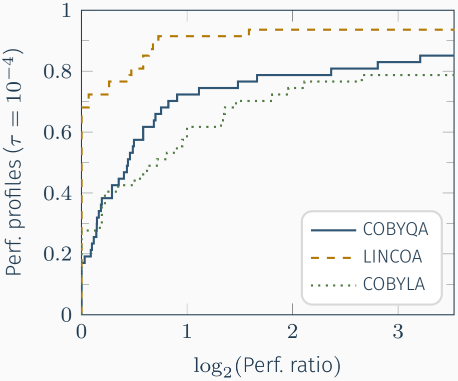
We compare two strategies for evaluating the Lagrange multipliers

- on nonlinearly constrained problems,
- comparing the default and the erroneous ($\lambda_k = 0$ and $\mu_k = 0$) ones.



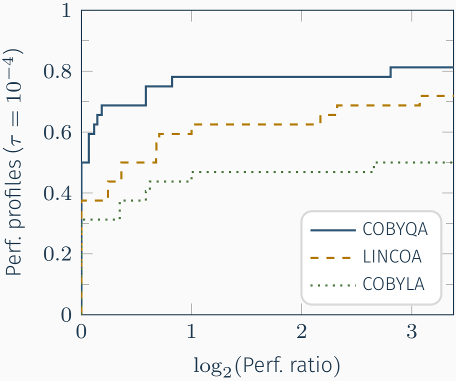
We compare COBYQA, LINCOA, and COBYLA

- on linearly constrained problems,



We compare COBYQA, LINCOA, and COBYLA

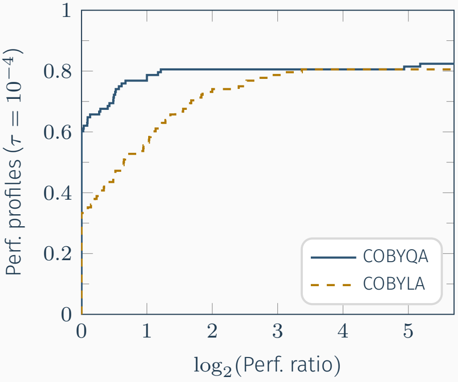
- on linearly constrained problems,
- with inviolable bounds.



Performance on nonlinearly constrained problems

We compare COBYQA and COBYLA

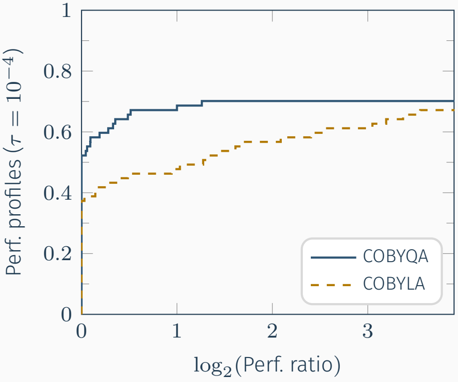
- on nonlinearly constrained problems,



Performance on nonlinearly constrained problems

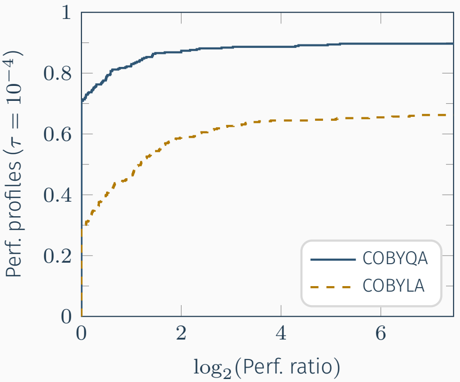
We compare COBYQA and COBYLA

- on nonlinearly constrained problems,
- with inviolable bounds.



Comparison with COBYLA

We compare COBYQA and COBYLA on all 388 problems.



Conclusion

Conclusion

We presented our new method COBYQA.

- It already received positive feedback from practitioners.
- It will soon be included in the Python packages PDFO and GEMSEO.

We established a new interpretation of the SQP subproblem.

- Does it provide new insights into manifold optimization?
- Can these insights help the theoretical analysis of COBYQA?

COBYQA

My thesis

```
$ pip install cobyqa
```

References I

- 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.* 134.1, pp. 223–257.
- 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, United Kingdom: Cambridge University Press, pp. 83–108.
- — (1997b). **"Recent progress in unconstrained nonlinear optimization without derivatives."** *Math. Program.* 79.1–3, pp. 397–414.
- — (1998). **"A derivative free optimization algorithm in practice."** In: *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*. St. Louis, MO, USA: AIAA, pp. 129–139.
- Conn, A. R., Scheinberg, K., and Vicente, L. N. (2008a). **"Geometry of interpolation sets in derivative free optimization."** *Math. Program.* 111.1–2, pp. 141–172.

References ii

- Conn, A. R., Scheinberg, K., and Vicente, L. N. (2008b). **"Geometry of sample sets in derivative-free optimization: polynomial regression and underdetermined interpolation."** *IMA J. Numer. Anal.* 28.4, 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.* 24.1, pp. 145–154.
- Garneau, M. L. (2015). **"Modelling of a solar thermal power plant for benchmarking blackbox optimization solvers."** MA thesis. Montreal, Quebec: Polytechnique Montréal.
- 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. Vol. 275. Mathematics and Its Applications. Dordrecht, Netherlands: Springer, pp. 51–67.
- — (2004). **"Least Frobenius norm updating of quadratic models that satisfy interpolation conditions."** *Math. Program.* 100.1, 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. Vol. 83. Nonconvex Optimization and Its Applications. Boston, MA, USA: Springer, pp. 255–297.
- Ragonneau, T. M. (2022). **“Model-Based Derivative-Free Optimization Methods and Software.”** PhD thesis. Hong Kong SAR, China: The Hong Kong Polytechnic University.
- Wild, S. M. (2008). **“MNH: a derivative-free optimization algorithm using minimal norm Hessians.”** In: *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.
- Zhang, Z. (2014). **“Sobolev seminorm of quadratic functions with applications to derivative-free optimization.”** *Math. Program.* 146.1–2, pp. 77–96.