

ONGOING WORK ON MCWAL

PIERRE BORIE

ABSTRACT. This informal document reflects the ongoing work and thinking on a algorithm for constrained nonlinear least squares. The current algorithm (rapper) name is MCWAL for Moindres Carrés With Augmented Lagrangian.

1. INTRODUCTION

We consider least squares problems subject to both nonlinear and linear constraints of the form

$$(1.1) \quad \begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} \|r(x)\|^2 \\ \text{s.t.} \quad & c(x) = 0 \\ & Ax = b \\ & \ell \leq x \leq u, \end{aligned}$$

where $r: \mathbb{R}^n \rightarrow \mathbb{R}^d$ and $c: \mathbb{R}^n \rightarrow \mathbb{R}^t$ are assumed to be nonlinear, potentially non convex, continuously differentiable functions, $\|\cdot\|$ denotes the euclidean norm, A is a full rank $m \times n$ matrix ($m \leq n$), $b \in \mathbb{R}^m$ and ℓ and u are vectors in \mathbb{R}^n . Without loss of generality, some components of the latter two vectors can be set to $\pm\infty$ for unbounded parameters. In the context of least squares problems, components r_i of the function r are often denoted as the residuals.

We will also refer to the linear constraints using the set notation ℓ

$$(1.2) \quad X = \{x \in \mathbb{R}^n \mid Ax = b, \ell \leq x \leq u\}.$$

1.1. Other notations and generalities on least squares. $\langle \cdot, \cdot \rangle$ is the canonical inner product, the one inducing the norm $\|\cdot\|$.

Rewriting the objective function of problem (1.1) as $f: x \mapsto \frac{1}{2} \|r(x)\|^2$, one has:

$$(1.3a) \quad \nabla f(x) = J(x)^T r(x)$$

$$(1.3b) \quad \nabla^2 f(x) = J(x)^T J(x) + S(x),$$

where $J(x) = \left[\frac{\partial r_i}{\partial x_j} \right]_{(i,j)}$ is the Jacobian matrix of the residuals and the second component of the Hessian $S(x) = \sum_{i=1}^d r_i(x) \nabla^2 r_i(x)$. The latter is expensive to compute and store, since it requires d computations of $n \times n$ matrices. Hence, this component of the Hessian is the one to be approximated.

When considering iterative methods for solving problem (1.1), k will, if not mentioned otherwise, refer to the iteration number. In order to simplify notations, quantities relative to a given iteration will be noted with the iteration number as an index, such as x_k for the iterate, r_k for $r(x_k)$, J_k for $J(x_k)$ etc.

We now address is a quick review of the three most popular classes of approximations. For a comprehensive review of these reviews, we refer the reader to the chapter 10 of [2].

The **Gauss–Newton** approximation sets $S(x)$ to the zero matrix. It is the cheapest to compute, since the Jacobian is necessary to evaluate the gradient and works well in practice for zero residuals problems [2]. When solving problem (1.1) using an iterative method, this approximation amounts to linearizing the residuals function within the norm.

Next, the **Levenberg–Marquardt** (LM) method [4, 5] sets $S(x)$ to a multiple of the identity matrix σI where σ is a regularization parameter. It is still cheap to compute but only requires updating the regularization parameter throughout the iterative process. This method can be seen as the early stage of the trust region methods [1]. It works well in practice on zero and small residuals problems.

Finally, one can compute an approximation of $\nabla^2 f(x)$ in a similar pattern as in **quasi-Newton** methods [6, Chapter 6] but targeted on the second order components. It has a higher computational cost than the previous two but is more accurate on large residuals problems. In [3], the authors exploit this approach in a adaptative scheme, where an estimation of the curvature is used to decide whether or not the quasi-Newton approximation is worth to use compared to the Gauss-Newton one.

It is important to bear in mind that choosing between a "cheap" approximation and a quasi-Newton type one implies making compromises. Depending on the initialization, the quasi-Newton approximation will take some iterations to be good and the accuracy will not be there when most needed, i.e. at the starting point potentially far from the solution, and will match the Gauss-Newton close to the solution on small residuals problems. In other words, the quasi-Newton is not accurate enough when most needed and very accurate when a way cheaper alternative does the same job.

REFERENCES

- [1] A.R. Conn, N.I.M Gould, and Ph.L. Toint. *Trust Region Methods*. SIAM: Society of Industrial and Applied Mathematics, Philadelphia, PA, USA, 2000. doi: 10.1137/1.9780898719857.
- [2] J.E. Dennis Jr and R.B. Schnabel. *Numerical Methods for Unconstrained optimization and Nonlinear Equations*. Classics in Applied Mathematics. SIAM, Philadelphia, PA, USA, 1996. doi: 10.1137/1.9781611971200.
- [3] J.E. Dennis Jr, D.M. Gay, and R.E. Walsh. An adaptive nonlinear least-squares algorithm. *ACM Transactions on Mathematical Software*, 7(3):348–368, 1981. doi: 10.1145/355958.355965.
- [4] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2:164–168, 1944.
- [5] D.W. Marquardt. An algorithm for least squares estimation of non-linear parameters. *SIAM Journal*, 11:431–441, 1963.
- [6] J. Nocedal and S.J. Wright. *Numerical Optimization*. Springer series in Operation Reasearch and Financial Engineering. Springer, New York, NY, USA, seconde edition, 2006. doi: 10.1007/978-0-387-40065-5.