On the Computation of Convergence Regions for Sequential Nonlinear Programming Problems

Ivan Mykhaliuk, Kai Schäfer, Kathrin Flaßkamp, and Christof Büskens
1 Optimization and Optimal Control, Center for Industrial Mathematics (ZTeM), University of Bremen, Bibliothekstraße 5, 28359 Bremen, Germany
2 Systems Modelling and Simulation, Saarland University, Campus, 66123 Saarbrücken, Germany

In this work, we formulate and solve the problem of finding the ball of maximum radius around a local minimum of a nonlinear optimization problem, which is invariant with respect to the gradient descent method. This problem arises in the context of solving sequences of nonlinear optimization problems, in which one usually strives to converge to qualitatively similar solutions. We illustrate our idea with an example of a nonlinear function of two variables.

© 2021 The Authors Proceedings in Applied Mathematics & Mechanics published by Wiley-VCH GmbH

1 Introduction

In many real-world applications, one has to solve a sequence of related nonlinear optimization problems (NLPs). For instance, a certain NLP might undergo perturbations that produce a set of “shifted” problems to be solved consecutively. In that case, one strives to avoid “jumps” from one local minimizer to another which might occur in such a scheme. Typical globalization techniques such as Armijo or Wolfe conditions [1] are not sufficient to guarantee the convergence to any fixed minimizer, hence, a different approach needs to be chosen.

As a more formal setting, consider the parameterized NLP

\[
\min \; f(x + s), \; x \in \mathbb{R}^n,\tag{1}
\]

where \( s \in \mathbb{R}^n \) is a perturbation which corresponds to a “shift” of the original problem, and let \( x^* \) be a local minimizer in the case \( s = 0 \). Further, let \( \{s^{(i)}\}_{i=1}^N \subseteq \mathbb{R}^n \) be a sequence of perturbations and consider the corresponding sequence of NLPs, in which the solution of the previous problem is the initial guess for the current one:

\[
\min \; f(x + s^{(i)}), \; x \in \mathbb{R}^n, \; x^{(i)} = x^{*^{(i-1)}}, \; i = 1, 2, \ldots, \tag{2}
\]

and \( x^{(0)} = x^* \). We assume that for each NLP (2) the same local optimization method is used.

For the process which arises from solving Eq. (2), our question is the following: what is the maximum perturbation magnitude \( d \) (in the sense of the Euclidean norm \( \| \cdot \|_2 \) ) such that if \( \| s^{(i)} \|_2 < d \) for all \( i = 1, 2, \ldots \), then Eq. (2) yields the sequence of solutions \( \{x^* - s^{(i)}\}_{i=1}^N \). In other words, we want to guarantee the convergence to the same (suitably shifted) minimizer \( x^* \).

To answer the question above, we need to find the maximum radius \( r^* \) of the ball \( B(x^*; r^*) := \{x \in \mathbb{R}^n : \|x - x^*\|_2 < r^*\} \) such that for all \( x_0 \in B(x^*; r^*) \), the sequence of points \( \{x_k\}_k \) produced by the local method starting with \( x_0 \) converges to \( x^* \). This problem can be formulated as follows:

\[
\max \; r \; \text{s.t.} \int_{B(0,r)} g(x; x^*) \, dx = 0, \; r \geq 0, \tag{3}
\]

where \( g \) is a certain non-negative function related to the chosen optimization algorithm.

2 Application to Gradient Descent Method

Let us now consider a particular application of Problem (3). Let \( f : \mathbb{R}^2 \to \mathbb{R} \) be a target function, and \( x^* \) be a local minimizer. As an optimization algorithm, we choose the gradient descent method (GDM)

\[
x_{k+1} = x_k - \gamma_k \nabla f(x_k), \; x_0 \in \mathbb{R}^2, \; \gamma_k > 0 \; \forall k = 1, 2, \ldots \tag{4}
\]

In this case, we answer a slightly weaker question, namely, to find the ball of the maximum radius \( r^* \) such that the points \( \{x_k\}_k \) produced by GDM (4) satisfy \( x_k \in B(x^*; r^*) \) for all \( k = 1, 2, \ldots \) whenever \( x_0 \in B(x^*; r^*) \).

* Corresponding author: e-mail ivamyk@uni-bremen.de
Fix the angle $\bar{\alpha} \in \left(0, \frac{\pi}{2}\right)$. For $x \neq x^*$ we require that the angle $\alpha$ between $\nabla f(x) := x^* x$ and $\nabla f(x) \neq 0$ be at most $\bar{\alpha}$ (see Fig. 1). Hence, there exists $t > 0$ such that $p(t) := x - t \nabla f(x)$ lies on the circle of radius $\|\nabla f(x)\|$ centered at $x^*$, and for each $s \in (0, t)$ we have that $\|x^* - p(s)\|_2 < \|x^* - x\|_2$. Formally, for the variation $0 \neq v \in \mathbb{R}^n$ we require that
cos (\alpha) := \frac{\nabla f(x^* + v)}{\|\nabla f(x^* + v)\|_2} \geq \cos (\bar{\alpha}).$

Hence, setting $h(x; x^*, \bar{\alpha}) := \nabla f(x^* + x) \cdot x - \|\nabla f(x^* + x)\|_2 \|x\|_2 \cos (\bar{\alpha})$ and
g (x; x^*, \bar{\alpha}) := |h (x; x^*, \bar{\alpha})| - h (x; x^*, \bar{\alpha})$

in Problem (3), we can find a radius $r^*$ (which, in general, can be infinite) such that the condition $\alpha \leq \bar{\alpha}$ or $\nabla f(x) = 0$ holds uniformly in $B(x^*; r^*)$. Furthermore, for the finite $r^*$ the inequality
$\gamma_k \|\nabla f(x_k)\|_2 \leq r^* \cos (\bar{\alpha}) \ \forall k = 1, 2, \ldots$

ensures that all the points $\{x_k\}_k$ produced by GDM (4) satisfy $x_k \in B(x^*; r^*)$ for all $k = 1, 2, \ldots$ provided $x_0 \in B(x^*; r^*)$.

To illustrate this idea, consider the function
$f(x, y) = (1 - x^2 - y^2) e^{-0.5(x^2+y^2)} - 1.5 e^{-3.2(x-1.5)^2-12(y-0.7)^2}$

with two local minimizers $M_1 \approx (1.514, 0.702)$ and $M_2 \approx (0.001, 1.879)$. Fig. 2 demonstrates the solutions $r^*_1 \approx 0.62$ and $r^*_2 \approx 1.1$ of Problem (3) with the function $g$ given by Eq. (5) and $\bar{\alpha} = \frac{\pi}{2} - 0.1$ for the minimizers $M_1$ and $M_2$, respectively. Both problems are solved by Monte Carlo based algorithm (see e.g. [2]). Fig. 2 also shows the iterative process of GDM for two different initial guesses. As one can see, the disks $B(M_1; r^*_1)$ and $B(M_2; r^*_2)$ are indeed maximum, i.e. choosing the initial guess even a little bit outside of $B(M_2; r^*_2)$ causes the corresponding GDM scheme to converge to a different solution.

### 3 Future Work

In this work, we formulate and analyze a model of sequential NLPs and show how the maximal radius of the ball which is invariant with respect to GDM can be obtained. In the future, we aim to provide a link to exemplary real-world applications such as model predictive control, in which consecutive optimal control problems are typically similar to each other [3]. On the other hand, the above-mentioned ideas have to be extended to methods other than GDM, in particular, the constrained optimization algorithms. Finally, a clear connection to the field of robust optimization [4] needs to be established.

**Acknowledgements** 1. Mykhailiuk acknowledges funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project number 281474342/GRK2224/1.
Open access funding enabled and organized by Projekt DEAL.

**References**

[1] J. Nocedal and S. J. Wright, Numerical Optimization, second edition (Springer, New York, NY, USA, 2006).

[2] N. Thomopoulous, Essentials of Monte Carlo Simulation (Springer New York, 2012).

[3] C. Terrell, Predictions in Time Series Using Regression Models (EDTECH, 2019).

[4] D. Bertsimas, D. Brown, and C. Caramanis, SIAM Review 53(10), 464–501 (2010).