A trust-region method with models based in support vector machines regression

Adriano Verdério

PPGMA - Programa de Pós-graduação em Matemática - UFPR
CEM - Centro de Engenharias da Mobilidade - UFSC
Campus Joinville
89218-000, Joinville, SC
E-mail: adriano.verderio@ufsc.br

Katya Scheinberg

Department of Industrial and Systems Engineering - Lehigh University
Harold S. Mohler Laboratory, 200 West Packer Avenue
Bethlehem, PA 18015-1582
E-mail: katyas@lehigh.edu

Elizabeth W. Karas, **Lucas G. Pedroso**.

Departamento de Matemática - UFPR
Centro Politécnico - Jardim das Américas
81531-980 Caixa Postal 19081, Curitiba, PR
E-mail: ewkaras@ufpr.br, lucaspedroso@ufpr.br.

1 Introduction

Consider the nonlinear programming problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega, \end{array} \tag{1}$$

where $\Omega \subset \mathbb{R}^n$ is a nonempty closed convex set and $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function.

Conejo et al. in [1] propose a globally convergent trust region algorithm to solve the problem (1). The algorithm allows great freedom in the construction of sub-problems and resolutions, and the case of interest is when the derivatives of the objective function are not available even though it is supposed to exist. Problems of this type are known as optimization problems without derivatives, that can arise when we want to optimize a function that come from a simulation, for example.

A trust-region algorithm define in each iteration a objective function model and a region where the model well approximates the function. If an adequate model of the objective function is found within the trust region then the model minimizer is the new iterate and the region can be expanded. Conversely, if the approximation is poor then the iterate is rejected and the region can be contracted to found a new minimizer.

In the algorithm propose in [1] any method can be used to build the model provided that is a good objective function approximation. Our goal is show that models built by support vector regression can be used and preserve the global convergence.

2 Support vector machines regression

Support vector machines (SVM) are a class of algorithms for Machine Learning motivated by results of Statistical Learning Theory. At the beginning, were used for pattern classification and subsequently extended to regression functions. In a sense it is a generalization of the usual regression techniques. Our goal is to use them to approximate functions which we have limited knowledge.

Suppose that we have the set $X = \{x^1, x^2, \ldots, x^m\}$ and we know the function in these points, an allowed error ε a constant C. To find a linear model $m(x) = w^\top x + b$ with support vector machines regression (SVMR) we need to solve the convex quadratic problem

$$\min \frac{1}{2}\|w\| + C\frac{1}{m}\sum_{i=1}^{m} \max\{0, |f(x^i) - w^\top x^i - b - \varepsilon|\},$$

which is equivalent to solving the problem

$$\begin{array}{ll} \min & \frac{1}{2}\|w\|^2 + C\frac{1}{m}\sum_{i=0}^{m}(\xi_i + \xi_i') \\ \text{s.t.} & w^\top x^i + b - f(x^i) \leq \varepsilon + \xi_i \\ & f(x^i) - w\top x_i + b \leq \varepsilon + \xi_i' \\ & \xi_i, \xi_i' \geq 0, \end{array}.$$

Usually we solve de dual problem since is a simpler one. Consider

$$P = \begin{bmatrix} (x^1)^\top \\ (x^2)^\top \\ \vdots \\ (x^m)^\top \end{bmatrix}, \quad Q = \begin{bmatrix} PP^\top & -PP^\top \\ -PP^\top & PP^\top \end{bmatrix}, \quad z = \begin{bmatrix} \alpha \\ \gamma \end{bmatrix},$$

$$v = \begin{bmatrix} -f(P) + \varepsilon e \\ f(P) - \varepsilon e \end{bmatrix} \quad \text{e} \quad A = [-e^\top, e^\top],$$

the dual problem is

$$\begin{array}{ll} \min & z^\top Q z + v^\top z \\ \text{s.t.} & Az = 0 \\ & 0 \leq z \leq C. \end{array}.$$

By the dual solution, we write $w = P^\top(\alpha - \gamma)$ and the linear model is $m(x) = w^\top x + b$, where b is computed by the KKT conditions as $b = f(x^i) - w^\top x^i - \varepsilon$ for any i with $0 < \alpha_i < C$.

To build a quadratic model we use a map $\varphi : \mathbb{R}^n \to \mathbb{R}^q$ to change to a higher dimensional space where the model is $q(x) = w^\top \varphi(x) + b$, it means that again we need to solve a convex quadratic problem.

With the solution of the convex quadratic problem we build the model and use it in the trust-region method propose in [1]. Under certain conditions we want to guarantee that the method still globally convergent when we use the support vector machines regression to build the model.

References

[1] P. D. Conejo, E. W. Karas, L. G. Pedroso, A. A. Ribeiro and M. Sachine. Global convergence of trust-region algorithms for convex minimization without derivatives. *Applied Mathematics and Computation*, 220:324–330, 2013.

[2] A. R. Conn, N. I. Gould, e Ph. L. Toint. *Trust-Region Methods.* Society for Industrial and Applied Mathematics, Philadelphia, 2000.

[3] A. R. Conn, K. Scheinberg, e L. N. Vicente. *Intoduction to Derivatife-Free Optimization.* Society for Industrial and Applied Mathematics, Philadelphia, 2009.

[4] J. Nocedal e S. J. Wright. *Numerical Optimization.* Springer Series in Operations Research. Springer-Verlag, New York, 1999.

[5] B. Schölkopf e A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond.* The MIT Press, Cambridge, 2002.

[6] V. N. Vapnik. *The Nature of Statistical Learning Theory.* Springer-Verlag, New York, 2000.