Applied Numerical Mathematics 60 (2010) 208–223

Contents lists available at ScienceDirect

Applied Numerical Mathematics

www.elsevier.com/locate/apnum

An entropy regularization technique for minimizing a sum of Tchebycheff norms

Suyan He [a,*], Yuxi Jiang [b], Shaohua Pan [c,1]

[a] School of Software, Dalian University of Foreign Languages, Dalian 116044, China
[b] School of Management, Dalian Jiaotong University, Dalian 116028, China
[c] Department of Mathematics, South China University of Technology, Guangzhou 510640, China

ARTICLE INFO

Article history:
Received 2 March 2009
Received in revised form 26 October 2009
Accepted 9 November 2009
Available online 22 November 2009

Keywords:
Sum of Tchebycheff norms
Entropy regularization
Smoothing method
Primal–dual trajectory
Quadratic convergence

ABSTRACT

In this paper, we consider the problem of minimizing a sum of Tchebycheff norms $\Phi(x)=\sum_{i=1}^{m}\|b_i-A_i^Tx\|_\infty$, where $A_i\in\mathbb{R}^{n\times d}$ and $b_i\in\mathbb{R}^d$. We derive a smooth approximation of $\Phi(x)$ by the entropy regularization technique, and convert the problem into a parametric family of strictly convex minimization. It turns out that the minimizers of these problems generate a trajectory that will go to the primal–dual solution set of the original problem as the parameter tends to zero. By this, we propose a smoothing algorithm to compute an ϵ-optimal primal–dual solution pair. The algorithm is globally convergent and has a quadratic rate of convergence. Numerical results are reported for a path-following version of the algorithm and made comparisons with those yielded by the primal–dual path-following interior point algorithm, which indicate that the proposed algorithm can yield the solutions with favorable accuracy and is comparable with the interior point method in terms of CPU time for those problems with $m\gg\max\{n,d\}$.

1. Introduction

Let $b_1,\ldots,b_m\in\mathbb{R}^d$ be column vectors in the Euclidean d-space and $A_1,\ldots,A_m\in\mathbb{R}^{n\times d}$ are n-by-d matrices. The problem of minimizing a sum of Tchebycheff norms is

$$\min_{x\in\mathbb{R}^n}\sum_{i=1}^{m}\|b_i-A_i^Tx\|_\infty, \tag{1}$$

where $\|\cdot\|_\infty$ represents the *Tchebycheff norm* of a vector, defined by

$$\|u\|_\infty=\max\{|u_1|,|u_2|,\ldots,|u_d|\}\quad\forall u\in\mathbb{R}^d.$$

Clearly, $x=0$ is an optimal solution to (1) when all of the b_i are zero. Therefore, in the rest of this paper we assume that not all of the b_i are zero.

Note that (1) is a nondifferentiable convex programming problem, and the function

* Corresponding author.
E-mail addresses: suyanhe@sina.com (S. He), shhpan@scut.edu.cn (S. Pan).

[1] The author's work is supported by National Young Natural Science Foundation (No. 10901058) and Guangdong Natural Science Foundation (No. 9251802902000001).

0168-9274/$30.00 © 2009 IMACS. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.apnum.2009.11.005

$$\Phi(x) := \sum_{i=1}^{m} \left\| b_i - A_i^T x \right\|_\infty \tag{2}$$

is not differentiable at any point x where, for some $i = 1, 2, \ldots, m$, there are more than two indices j such that $|(b_i - A_i^T x)_j| = \max_{1 \leqslant l \leqslant d} |(b_i - A_i^T x)_l|$, with $(b_i - A_i^T x)_l$ denoting the lth component of $b_i - A_i^T x$. The nonsmooth problem arises in many applications, which induce the Tchebycheff norm multi-facility location problems [6,11] and the Tchebycheff norm Steiner minimal tree problems under a given topology [19,20]. When $d = 1$, it models the linear l_∞ problem; see [5] and references therein. Particularly, we see that the problem (1) arising from the practical applications often satisfies $m \gg \max\{d, n\}$.

Many practical algorithms have been designed to solve the problem of minimizing a sum of Euclidean norms (see, e.g., [1,2,4,14,16,17,19]), but few are given for the problem of minimizing a sum of Tchebycheff norms except the interior point methods in [20,21]. In this paper, we develop a smoothing method for problem (1). Specifically, we derive a smooth approximation $\Phi_\tau(x)$ of $\Phi(x)$ by using the entropy regularization technique [10,13], and then transform (1) to the solution a family of strictly convex minimization problems with $\Phi_\tau(x)$ being the objective. The minimizers of these problems generate a trajectory going to the primal–dual solution set as the regularizing parameter τ approaches to 0. By this, we propose a noninterior point primal algorithm by solving a single approximation problem with the classical Newton method, and show that the algorithm can locate a point on the solution trajectory, i.e. find an ϵ-optimal primal–dual solution pair of problem (1), at a quadratic rate of convergence.

Note that (1) can be reformulated as the following linear programming problem:

$$\begin{aligned} &\min \sum_{i=1}^{m} y_i \\ &\text{s.t.} \quad A_i^T x - y_i e_d \leqslant b_i, \quad i = 1, 2, \ldots, m, \\ &\qquad\ A_i^T x + y_i e_d \geqslant b_i, \quad i = 1, 2, \ldots, m, \end{aligned} \tag{3}$$

where $y_i \in \mathbb{R}$ for $i = 1, 2, \ldots, m$ and e_d represents the column vector of ones in $\mathbb{R}^d$. Hence, the successful interior point methods can be applied for the problem. Following this line, Zhang [21] developed a primal–dual path-following interior point method for a special case of (1) where $m = 1$. Compared with the primal–dual interior point method, our smoothing algorithm is a primal one but can seek an ϵ-optimal primal dual solution pair at a satisfying convergence rate, i.e. the quadratic convergence rate. In addition, as will be shown in Section 4, the algorithm requires only one multiplication of $n \times n$ matrix and $n \times 1$ vector and one multiplication of $n \times md$ matrix and $md \times 1$ vector, except the factorization of an n-by-n positive definite matrix. This implies that our smoothing algorithm requires less computation work in each iteration, especially for those problems with $m \gg \max\{n, d\}$. Numerical comparisons in Section 5 also illustrate this fact.

The rest of this paper is organized as follows. Section 2 is devoted to the favorable properties of the smooth approximation $\Phi_\tau(x)$. Section 3 studies the behavior of the trajectory generated by the minimizer of the approximation problem with $\Phi_\tau(x)$ being the objective. In Section 4, we propose a smoothing algorithm by solving a single approximation problem, and establish the global and quadratic convergence results for the algorithm. In Section 5, we report numerical results for a path-following version of the smoothing method with solving Tchebycheff norm multi-facility location problems and Steiner minimal tree problems, and compare the numerical performance with that of the primal–dual path-following interior point algorithm [21] which was based on the KKT optimality conditions of (3). Finally, we conclude this paper.

Throughout this paper, e_d represents the column vector of ones in $\mathbb{R}^d$. For a given $x \in \mathbb{R}^n$, we denote the ith component of x by x_i or $(x)_i$, and the diagonal matrix with the components of x as diagonals by $\mathrm{diag}(x)$ or $\mathrm{diag}(x_1, \ldots, x_n)$. Given a twice continuously differentiable function $f(x)$, $\nabla_x f(x)$ and $\nabla^2_{xx} f(x)$ denote the gradient and Hessian matrix of f at x, respectively. To represent a large matrix with several small matrices, we use semicolons ";" for column concatenation and commas "," for row concatenation. This notation also applies to vectors. Given a finite number of square matrices $Q_1, \ldots, Q_n$, we denote the block diagonal matrix with these matrices as block diagonals by $\mathrm{diag}(Q_1, \ldots, Q_n)$ or by $\mathrm{diag}(Q_i,\ i = 1, 2, \ldots, n)$. The notation $\tau \downarrow 0$ denotes the case that a positive scalar τ approaches to 0. In addition, in the sequel, we write

$$A = [A_1\ A_2\ \ldots\ A_m] \in \mathbb{R}^{n \times (md)} \quad \text{and} \quad b = (b_1; b_2; \ldots; b_m) \in \mathbb{R}^{md},$$

and assume that at least one A_i for $i = 1, 2, \ldots, m$ has full row rank n.

2. Smooth approximation and its properties

In this section, we derive a smooth approximation of $\Phi(x)$ by the entropy regularization technique [10,13]. For this purpose, we first represent $\Phi(x)$ by a linear programming problem. Let $z_1, z_2, \ldots, z_m \in \mathbb{R}^d$. By the definition of dual norm [8], it follows that

$$\Phi(x) = \max\left\{ \sum_{i=1}^{m} z_i^T \left(b_i - A_i^T x\right):\ \|z_i\|_1 \leqslant 1,\ i = 1, 2, \ldots, m \right\}.$$