Barvinok's naive algorithm in Distance Geometry

Leo Liberti, Ky Vu

PII: S0167-6377(17)30682-X
DOI: https://doi.org/10.1016/j.orl.2018.06.006
Reference: OPERES 6376

To appear in: *Operations Research Letters*

Received date : 13 December 2017
Revised date : 25 June 2018
Accepted date : 25 June 2018

Please cite this article as: L. Liberti, K. Vu, Barvinok's naive algorithm in Distance Geometry, *Operations Research Letters* (2018), https://doi.org/10.1016/j.orl.2018.06.006

Barvinok's naive algorithm in Distance Geometry☆

Leo Liberti

CNRS LIX, Ecole Polytechnique, 91128 Palaiseau, France

Ky Vu

Chinese University of Hong Kong, P.R. China

Abstract

In 1997, A. Barvinok gave a probabilistic algorithm to derive a near-feasible solution of a quadratically (equation) constrained problem from its semidefinite relaxation. We generalize this algorithm to handle matrix variables instead of vectors, and to two-sided inequalities instead of equations. We derive a heuristic for the distance geometry problem, and showcase its computational performance on a set of instances related to protein conformation.

Keywords: distance geometry, concentration of measure, protein structure

---

1. Introduction

We consider the following

> Distance Geometry Problem (DGP). Given an integer $K > 0$ and a simple graph $G = (V, E, d)$ where $d : E \to \mathbb{R}_+$, decide whether there is a realization $x : V \to \mathbb{R}^K$ such that
>
> $$\forall \{i, j\} \in E \qquad \|x_i - x_j\|_2^2 = d_{ij}^2. \tag{1}$$

Let $n = |V|$ and $m = |E|$. The DGP is well known in the literature [1, 2], as it serves as a model for several applications (e.g. to sensor networks [3], molecular conformation [4, 5] and more [6]). A natural extension of the DGP is the interval version (denoted by *i*DGP), where $d : E \to \mathbb{IR}_+$ associates intervals (instead of scalars) to edges. This variant is the one often used in applications, since intervals naturally model data uncertainty and noise [7].

We adapt the *naive algorithm* proposed by A. Barvinok in [8, §5] to the DGP setting. Barvinok's main insight is that the solutions of a Semidefinite Programming (SDP) relaxation are "not too far" from the feasible set of the Quadratically Constrained Problem (QCP) which gives rise to the SDP relaxation: it suffices to factor the SDP solution and multiply it by a random vector having components sampled from a normal distribution. The very natural idea we propose is to use the naive algorithm first, then use this approximate solution as a starting point for a local Nonlinear Programming (NLP) solver deployed on the original QCP, hoping it will converge to a realization satisfying the feasibility of Eq. (1).

Barvinok's naive algorithm is a randomized algorithm based on the concentration of measure phenomenon. It applies to SDP relaxations of systems of quadratic equations, i.e. pure feasibility, equation-only QCPs having vector solutions, or, equivalently, $n \times 1$ matrix solutions. This algorithm cannot be natively applied to the DGP, since a realization x is naturally represented by an $n \times K$ matrix $x = (x_{ik} \mid i \le n, k \le K)$ the i-th row of which is the position vector of vertex $i \in V$ in $\mathbb{R}^K$. This matrix may in general have rank greater than one. In [8], Barvinok gives a proof sketch which only applies to $n \times 1$ matrices. Of course, an $n \times K$ matrix can also be represented as an $n' \times 1$ matrix where $n' = nK$, but this would entail a QCP (and hence also an SDP relaxation) with $nK \times nK$ data matrices, which is practically prohibitive to solve. Moreover, the naive algorithm cannot be applied to the *i*DGP since it is constrained by quadratic inequalities rather than equations.

We make two contributions in this paper, one theoretical and the other computational. (i) We propose a generalization of Barvinok's result [8] in two directions: the $n \times K$ case, and the case of QCP inequalities of the form $d^L \le x^\top Q x \le d^U$ (where x is $n \times K$). (ii) We establish the practical usefulness of Barvinok's result on a set of medium and large-scale DGP instances extracted from the Protein Data Bank (PDB) [9]. We also remark that our proof is detailed, and fills many gaps in Barvinok's "proof sketch" found in [8]. The rest of this paper is organized as follows. In Sect. 2 we present Barvinok's naive algorithm. In Sect. 3 we state and prove our generalization of Barvinok's concentration of measure result. In Sect. 5 we discuss our computational results.

☆We are grateful to A. Barvinok for helpful suggestions and discussions, as well as to an anonymous referee for helping us improve the paper. The first author (LL) has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement n. 764759 (ITN "MINOA").

Email addresses: `liberti@lix.polytechnique.fr` (Leo Liberti), `vukhacky@gmail.com` (Ky Vu)