COST TD1207 STSM 33826 Report

Leo Liberti
CNRS LIX, Ecole Polytechnique
liberti@lix.polytechnique.fr

April 28, 2016

Abstract

Summary of discussions between F. Tardella (host) and L. Liberti (guest) during an academic visit to Università di Roma "La Sapienza" funded by COST TD1207, in view of future joint works and visits.

1 Purpose of the STSM

The main purpose of the visit was to establish a scientific ground for discussing research topics in combinatorial and nonlinear optimization.

2 Description of the work carried out during the STSM

I gave a seminar titled "Quantile regression on large data sets — Random projections in linear programming", which focused on applying random projection techniques to dense linear programs arising in the well known statistical problem of quantile regression.

During the rest of my time in Rome, my host (Fabio Tardella) and I discussed several topics for further possible work on research ideas for combinatorial and nonlinear optimization. Some details about this discussion are summarized in the next session.

3 Description of the main results obtained

The purpose being the estbalishment of a common ground for future research, I feel the goal has been achieved. We now have a potential research agenda with enough broad topics to keep us occupied in the next few years.

3.1 Worst-case performance of random projections in optimization

In the framework of Johnson-Lindenstrauss type [6] random projections used in optimization [12], in particular in Linear Programming (LP) and Integer Programming (IP) [8], one has to transform the constraints of a Mathematical Program (MP) into a lower dimensional version thereof. For example, the constraints of an LP in standard form are $Ax = b$, where A is an $m \times n$ matrix: one has to find an

appropriate (sub-Gaussian) *random projector* T such that, in $TAx = Tb$ the matrix TA is $k \times n$ with $k \ll m$. In [8, 7] it is shown that if T is sampled componentwise from a normal distribution $N(0, \frac{1}{\sqrt{k}})$ and k is $O(\ln n)$ then

$$Ax = b \wedge x \geq 0 \text{ is feasible if and only if } TAx = Tb \wedge x \geq 0 \text{ is feasible} \qquad (*)$$

with probability approaching 1 exponentially fast as k increases. An important question in this setting is determining the worst case. With what probability is $(*)$ false, and, more importantly, how large can $\|Ax - b\|_2 - \min(x, 0)$ be?

3.2 Diagonally dominant programming

In a recent breakthrough, Amir Ali Ahmadi et al. [1] proposed a hierarchy of LPs which inner approximates the positive semidefinite (PSD) cone. By strong duality, this allows to both inner- and outer-approximate Semidefinite Programming (SDP) problems. Given that SDP solving technology is not as advanced as LP solving technology, this breakthrough promises a key to effective and strong LP relaxations of many problems. In [10] this was applied to control problems; in [4], the technique was applied to the Distance Geometry Problem (also see Sect. 3.3) — both are continuous nonlinear problems. But the technique should be tried and tested on a much broader spectrum of problems, including those of combinatorial type.

3.3 Distance geometry in protein conformation

The Distance Geometry Problem (DGP) is as follows. Given an integer $K > 0$ and a simple undirected non-negatively edge-weighted graph $G = (V, E, d)$ with $d : E \to \mathbb{R}_+$, establish whether there exists a *realization* function $x : V \to \mathbb{R}^K$ such that:

$$\forall \{i, j\} \in E \quad \|x_i - x_j\|_2 = d_{ij}.$$

This problem has a long history, and arises in many application fields [9], including protein conformation ($K = 3$), wireless sensor localization ($K = 2$), clock synchronization protocols ($K = 1$), coordination of unmanned underwater vehicles ($K = 3$) and others. In the application to protein conformation, the set V of vertices represents atoms in a protein backbone, and E represents the pairs for which the inter-atomic distance is known. For every such edge $e = \{i, j\} \in E$, d_{ij} is known fairly precisely if e represents an inter-atomic bond, or is a direct result of inter-atomic bonds and angles. On the other hand, d_{ij} is known as an interval if e is measured through Nuclear Magnetic Resonance (NMR). Various techniques are currently being tested in order to retain the efficiency of the mixed-combinatorial techniques used when the distances are known precisely in the case where some of them are known imprecisely, but the issue is still open.

3.3.1 The unassigned DGP

NMR experiments do not actually yield the value d_{ij} relative to given $i, j \in V$ — this is actually an outcome of a post-processing of the "raw output", which is a distribution over a set of triplets

$$(\texttt{atom\_type}_1, \texttt{atom\_type}_2, \texttt{distance\_value}).$$

Let A be the set of atom types (e.g. H, C, O, N,...) occurring in the molecule $\mathcal{M}$ under investigation. Then the output of an NMR experiment can be considered as a function $p : A \times A \times \mathbb{R}_+ \to [0, 1]$, where $p(a_1, a_2, \delta)$ is the probability that two atoms i, j with types a_1, a_2 respectively have distance δ. The Unassigned DGP (UDGP) consists in finding all graphs G which are compatible with the probability distribution p, i.e. the relative assignment of distance values to edges respects the distribution and

the atom types (note that problem statement is informal). As far as we know, NMR machinery uses a meta-heuristic (usually based on simulated annealing) to solve the UDGP. We also know that this methodology yields DGP instances which are usually infeasible, since around 10%-20% of the edges have the wrong distance [3]. Currently, we are aware of only one alternative method for solving the UDGP in the literature, namely TRIBOND [5], but this is also a heuristic. The issue here is to formally define the UDGP and then find exact solution methods for finding all graphs which are compatible with the NMR output.

3.4 Multiplicative weights update

Multiplicative weights update algorithms [2] are essentially "weighted Multi-Start" heuristics with an interesting twist: the overall performance of the algorithm is bounded by the performance of the best-performing weighted component. Although this is only a relative bound, it becomes absolute if one can bound the best-performing component using the problem structure (see e.g. [13]). How well does this perform for general optimization problems? The first results appear to be promising [11], but a much more detailed investigation is necessary.

4 Future collaboration with the host institution

We are looking into ways of funding a longer visit of mine to Fabio Tardella's department, possibly in 2017 or 2018. We are writing a research proposal for a call by Università di Roma "La Sapienza". If successful, I will be able to visit this department for up to 1 month.

References

[1] A. Ahmadi and G. Hall. Sum of squares basis pursuit with linear and second order cone programming. Technical Report 1510.01597v1, arXiv, 2015.

[2] S. Arora, E. Hazan, and S. Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory of Computing*, 8:121–164, 2012.

[3] B. Berger, J. Kleinberg, and T. Leighton. Reconstructing a three-dimensional model with arbitrary errors. *Journal of the ACM*, 46(2):212–235, 1999.

[4] G. Dias and L. Liberti. Diagonally dominant programming in distance geometry. In R. Mahjoub et al., editor, *International Symposium in Combinatorial Optimization*, LNCS, New York, to appear. Springer.

[5] P. Duxbury, L. Granlund, P. Juhas, and S. Billinge. The unassigned distance geometry problem. *Discrete Applied Mathematics*, to appear.

[6] W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In G. Hedlund, editor, *Conference in Modern Analysis and Probability*, volume 26 of *Contemporary Mathematics*, pages 189–206, Providence, 1984. American Mathematical Society.

[7] Vu Khac Ky, P.-L. Poirion, and L. Liberti. Gaussian random projections for euclidean membership problems. Technical Report 1509.00630v1[math.OC], arXiv, 2015.

[8] Vu Khac Ky, P.-L. Poirion, and L. Liberti. Using the johnson-lindenstrauss lemma in linear and integer programming. Technical Report 1507.00990v1[math.OC], arXiv, 2015.

[9] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino. Euclidean distance geometry and applications. *SIAM Review*, 56(1):3–69, 2014.

[10] A. Majumdar, A. Ahmadi, and R. Tedrake. Control and verification of high-dimensional systems with dsos and sdsos programming. In *Conference on Decision and Control*, volume 53, pages 394–401, Los Angeles, 2014. IEEE.

[11] L. Mencarelli, Y. Sahraoui, and L. Liberti. A multiplicative weights update algorithm for MINLP. *EURO Journal on Computational Optimization*, in revision.

[12] Mert Pilanci and Martin J Wainwright. Randomized sketches of convex programs with sharp guarantees. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 921–925. IEEE, 2014.

[13] S. Plotkin, D. Shmoys, and É. Tardos. Fast approximation algorithm for fractional packing and covering problems. *Mathematics of Operations Research*, 20:257–301, 1995.