Why does the Burer-Monteiro heuristic work so well?

Irène Waldspurger
Université Paris Dauphine
waldspurger@ceremade.dauphine.fr

Context

The broad topic of this internship proposal is non-convex optimization. It is a very active research domain. Indeed, while most optimization problems which naturally appear in applications are non-convex, these were, until recently, considered as almost impossible to solve with a “reasonable” computational cost. But in the last decade, it became apparent that many non-convex problems could actually be exactly solved, and with very simple algorithms.

The goal of the internship is to contribute to the theoretical understanding of this phenomenon, through the study of an important family of non-convex problems and its most widely-used solver.

If the internship goes well, it could be extended into a PhD, next fall.

Detailed subject

Optimization problems under consideration are semidefinite programs, that is, optimization problems of the following form:

\[
\begin{align*}
\text{minimize } & \; \text{Trace}(CX), \\
\text{for } & \; X \in \mathbb{R}^{n \times n} \text{ such that } A(X) = b, \\
& \; X \succeq 0,
\end{align*}
\]

where \( A : \mathbb{R}^{n \times n} \to \mathbb{R}^m \) is a linear map capturing \( m \) affine constraints on \( X \), and \( b \in \mathbb{R}^m \), \( C \in \mathbb{R}^{n \times n} \) are known. The non-convexity arises from the additional assumption that the solution has low rank.
These problems are important for applications, notably because they include high-quality approximations of NP-hard combinatorial optimization problems, like MaxCut.

To solve these problems, the most widely-used approach is the Burer-Monteiro heuristic. If $X^*$ is the desired solution, and $p$ is any integer larger than $\text{rank}(X^*)$, there must exist a matrix $V^* \in \mathbb{R}^{n \times p}$ such that

$$X^* = V^*(V^*)^T.$$ 

This observation allows to rewrite the semidefinite program in factorized form

$$\begin{align*}
\text{minimize} & \quad \text{Trace}(CVV^T), \\
\text{for} & \quad V \in \mathbb{R}^{n \times p} \text{ such that } A(VV^T) = b.
\end{align*}$$ 

The factorized form is more appealing than the initial one, from a computational point of view, because the unknown $V$ has only $np$ coefficients, which is usually much less than the $n^2$ coefficients of $X$. To solve it, we can apply standard local optimization algorithms, like (Riemannian) gradient descent.

In principle, because of non-convexity, these local optimization algorithms can get stuck at a non-optimal critical point instead of converging to the solution. However, empirically, it seems that non-convexity is only an issue when $p$ is very small (equal to $\text{rank}(X^*)$ or slightly larger); for other values of $p$, algorithms do not seem to encounter critical points. A rigorous explanation for this observation has been proposed in [Boumal, Voroninski, and Bandeira, 2020], but only for large values of $p$.

The goal of the internship is to (partially, at least) understand the regime where $p$ is only moderately large, as would typically happen in practice. A first step is to study random semidefinite programs and determine which properties control the existence or non-existence of critical points. Preliminary numerical experiments hint at a link between the existence of critical points and some form of instability in the semidefinite problem, but this requires more careful examination.

Bibliography
