AN APPROXIMATION ALGORITHM FOR THE BIPARTITE GROTHENDIECK PROBLEM

ALEXANDER IRIZA

JOINT WORK WITH AFONSO BANDEIRA & AMIT SINGER

1. Overview

The Grothendieck problem over the orthogonal group is given by the following optimization problem:

$$\max_{O_1,\ldots,O_n \in \mathcal{O}(d)} \operatorname{Tr}\left(\sum_{i,j=1}^{n} O_i O_j^T C_{ij}\right) \tag{1.1}$$

where C_{ij} is the ij^{th} $d \times d$ block of a symmetric matrix $C \in \mathbb{R}^{nd \times nd}$ and $\mathcal{O}(d)$ is the d-dimensional orthogonal group [2].

This problem is NP-hard, so there is no known algorithm to determine the solution in polynomial time. Instead we seek to compute an approximate solution in polynomial time. One strategy for accomplishing this, first pioneered by Goemans & Williamson in Ref. [3], is to employ a semidefinite relaxation. This technique consists of three steps: first construct a related problem that is a semidefinite program (and hence can be solved in polynomial time), then solve this problem, and finally develop a polynomial time procedure to round this solution back into a valid input to the original problem that achieves some fraction of the optimal objective value.

Ref. [2] considers the special case of Problem 1.1 called the little Grothendieck problem, in which the additional assumption that C is positive semidefinite is imposed. This case encodes several important optimization problems, such as the Procrustes problem and Global Registration. Ref. [2] develops a rounding procedure that produces a randomized solution whose expected objective value is at least a positive constant (the approximation ratio) times the objective value of the original problem. Their approximation ratio for all values of d exceeds the previous best known result of $\frac{1}{2\sqrt{2}}$ from Ref. [4], and their algorithm is much simpler and more efficient.

However in some applications (e.g. Orthogonal Synchronization) it is not appropriate to assume that $C \succeq 0$, so we would like to develop constant factor approximation ratios in other cases as well. In this paper we consider the following case of Problem 1.1, called the bipartite Grothendieck problem:

$$\max_{O_1,\ldots,O_n,U_1,\ldots,U_n \in \mathcal{O}(d)} \operatorname{Tr}\left(\sum_{i,j=1}^{n} O_i U_j^T C_{ij}\right) \tag{1.2}$$

If the matrix C is bipartite (i.e. is the adjacency matrix of a bipartite graph), then Problem 1.1 can be written in this form with the appropriate renaming of variables, so Problem 1.2 is a special case of Problem 1.1. One potential method for converting a problem of interest into bipartite form is to force C to be bipartite by disregarding up to half of the data.

In fact, we consider the following generalization of Problem 1.2 (for $d \leq r$):

$$\max_{O_1,\dots,O_n,U_1,\dots,U_n \in \mathcal{O}(d,r)} \operatorname{Tr}\left(\sum_{i,j=1}^{n} O_i U_j^T C_{ij}\right) \tag{1.3}$$

where $\mathcal{O}(d,r) = \left\{O \in \mathbb{R}^{d\times r} \,|\, OO^T = I_{d\times d}\right\}$ is the Stiefel manifold. This can encode the Common-Lines problem from Cryo-electron microscopy (see Ref. [2] for a discussion).

In Section 2 we develop a (randomized) polynomial-time rounding scheme for this problem that achieves a constant factor approximation ratio. In Section 3 we consider a specific instance of this rounding scheme and calculate the corresponding approximation ratio. We find that our ratio exceeds the ratio of $\dfrac{1}{2\sqrt{2}}$ from Ref. [4] (which is applicable to this problem) in some cases. Furthermore our procedure is simpler and more efficient than their algorithm. Finally in Section 4 we discuss future work.

2. Constant-Factor Approximation Ratio

We begin by posing the following relaxation of Problem 1.3:

$$\max_{V_1,\dots,V_n,W_1,\dots,W_n \in \mathcal{O}(d,2nd)} \operatorname{Tr}\left(\sum_{i,j=1}^{n} V_i W_j^T C_{ij}\right) \tag{2.1}$$

Letting $Y = [V_1^T \dots V_n^T \, W_1^T \, \dots W_n^T]$, $X = Y^T Y$, and $D = \begin{bmatrix} 0 & C \\ C & 0 \end{bmatrix}$, we see that Problem 2.1 is equivalent to

$$\max_{\substack{X \in \mathbb{R}^{2nd\times 2nd} \\ X \succeq 0 \\ X_{ii} = I_{d\times d}}} \operatorname{Tr}(DX) \tag{2.2}$$

so it is a semidefinite program. By Ref. [5], any semidefinite program can be solved with arbitrarily small error in polynomial time. Let ω_{SDP} denote the optimal value of this problem.

In the following definitions and results we explain our rounding scheme and approximation ratio. Our technique is a generalization of the one used by Alon and Naor in Section 4 of Ref. [1] for the case of $d = 1$.

Definition 2.1. An orthogonal rounding procedure is a map J that assigns to each $V \in \mathcal{O}(d, 2nd)$ a random matrix $J(V) \in \mathbb{R}^{d\times r}$ such that the following three properties are satisfied:

(1) $\mathbb{E}\left[J(V)J(W)^T\right] = VW^T$ for all $V, W \in \mathbb{R}^{d\times r}$.

(2) There exists $M > 0$ such that $\gamma(M,d,r) = 2\displaystyle\int_M^\infty \mathbb{P}\left\{\|J(V)\|_\infty \geq t\right\}(t-M)dt < \frac{1}{4}$ (where $\|\cdot\|_\infty$ denotes the spectral norm).

(3) J is computable in polynomial time in the size of the input.

Definition 2.2. For an orthogonal rounding procedure J and a real number $M > 0$, the truncated orthogonal rounding procedure J^M is defined by

$$J^M(V) = \begin{cases} \dfrac{J(V)}{M} & \|J(V)\|_\infty \leq M \\ \\ \dfrac{J(V)}{\|J(V)\|_\infty} & \|J(V)\|_\infty > M \end{cases}$$

Before turning to the main result (Theorem 2.5), we first show the following two lemmas:

Lemma 2.3. *For $V \in \mathcal{O}(d, 2nd)$,*

$$\mathbb{E}\left[\left\|J(V) - MJ^M(V)\right\|_\infty^2\right] = \gamma(M, d, r).$$

Proof.

$$\begin{aligned}
\mathbb{E}\left[\left\|J(V) - MJ^M(V)\right\|_\infty^2\right] &= \mathbb{E}\left[\left(\left(\|J(V)\|_\infty - M\right)_+\right)^2\right] \\
&= \int_0^\infty \mathbb{P}\left\{\left(\left(\|J(V)\|_\infty - M\right)_+\right)^2 \geq u\right\} du \\
&= \int_0^\infty \mathbb{P}\left\{\|J(V)\|_\infty \geq M + \sqrt{u}\right\} du \\
&= 2\int_M^\infty \mathbb{P}\left\{\|J(V)\|_\infty \geq t\right\}(t - M)dt \\
&= \gamma(M, d, r)
\end{aligned}$$

$\square$

Lemma 2.4. *Suppose that $A_1, \ldots, A_n, B_1, \ldots, B_n \in \mathbb{R}^{d\times r}$ are random variables such that*

$$\mathbb{E}\left[A_i A_i^T\right] \preceq \alpha I_{d\times d} \quad \textit{and} \quad \mathbb{E}\left[B_j B_j^T\right] \preceq \beta I_{d\times d}.$$

Then

$$\mathbb{E}\left[\operatorname{Tr}\left(\sum_{i,j=1}^n A_i B_j^T C_{ij}\right)\right] \leq \sqrt{\alpha\beta}\,\omega_{SDP}.$$

Proof. Let $\tilde{A}_i = \dfrac{1}{\sqrt{\alpha}}A_i$, $\tilde{B}_j = \dfrac{1}{\sqrt{\beta}}B_j$, $G = [\tilde{A}_1^T \ldots \tilde{A}_n^T\, \tilde{B}_1^T \ldots \tilde{B}_n^T]$, $H = G^T G$, and

$$\omega' = \mathbb{E}\left[\operatorname{Tr}\left(\sum_{i,j=1}^n A_i B_j^T C_{ij}\right)\right].$$

Using the notation of Problem 2.2, we have

$$\omega' = \sqrt{\alpha\beta}\operatorname{Tr}\left(D \cdot \mathbb{E}\left[H\right]\right).$$

Next we claim that $(\mathbb{E}\left[H\right])_{ii} \preceq I_{d\times d}$ and $\mathbb{E}\left[H\right] \succeq 0$.

First, if $i \leq nd$ then $(\mathbb{E}[H])_{ii} = \mathbb{E}\left[H_{ii}\right] = \mathbb{E}\left[\tilde{A}_i \tilde{A}_i^T\right] \preceq I_{d\times d}$ by the assumption that $\mathbb{E}\left[A_i A_i^T\right] \preceq \alpha I_{d\times d}$, and if $nd < i \leq 2nd$ then the same result holds by the assumption on the B_j's. Also, $\mathbb{E}[H] \succeq 0$ because $H \succeq 0$ always and the set of positive semidefinite matrices is convex.

Now recalling Problem 2.2, we see that

$$\omega_{SDP} = \max_{\substack{X\in\mathbb{R}^{2nd\times 2nd}\\ X\succeq 0\\ X_{ii}=I_{d\times d}}} \operatorname{Tr}(DX)$$

Because $D_{ii} = 0_{d\times d}$, this may be rewritten as

$$\omega_{SDP} = \max_{\substack{X\in\mathbb{R}^{2nd\times 2nd}\\ X\succeq 0\\ X_{ii}\preceq I_{d\times d}}} \operatorname{Tr}(DX)$$

But we have seen that $\mathbb{E}[H]$ is a valid input to this problem with objective value $\dfrac{\omega'}{\sqrt{\alpha\beta}}$, so

$$\frac{\omega'}{\sqrt{\alpha\beta}} \leq \omega_{SDP}.$$

□

Theorem 2.5. *Let $V_1, \ldots, V_n, W_1, \ldots, W_n \in \mathcal{O}(d, 2nd)$ constitute a solution to Problem 2.1, and let ω_{SDP} denote the optimal value of the problem. Also let J be an orthogonal rounding procedure and $M > 0$, and let*

$$\beta(M, d, r) = \frac{1 - 2\sqrt{\gamma(M, d, r)}}{M^2}$$

where $\gamma(M, d, r)$ is as in Definition 2.1.

Then

$$\mathbb{E}\left[\operatorname{Tr}\left(\sum_{i,j=1}^{n} J^M(V_i) J^M(W_j)^T C_{ij}\right)\right] \geq \beta(M, d, r)\, \omega_{SDP}$$

This can be optimized by taking $\beta(d, r) = \max_{M>0} \beta(M, d, r)$.

Because all singular values of $J^M(V_i), J^M(W_j)$ are at most 1 and the objective function is linear in these variables, they can be transformed to elements of $\mathcal{O}(d, r)$ without decreasing the objective value. Also, condition (2) from Definition 2.1 ensures that $\beta(d, r) > 0$. Therefore this provides a polynomial-time constant factor approximation algorithm for Problem 1.3.

Proof. By condition (1) of Definition 2.1, we see that

$$\begin{aligned}
\omega_{SDP} &= \operatorname{Tr}\left(\sum_{i,j=1}^{n} V_i W_j^T C_{ij}\right) \\
&= \mathbb{E}\left[\operatorname{Tr}\left(\sum_{i,j=1}^{n} J(V_i)J(W_j)^T C_{ij}\right)\right] \\
&= M^2\mathbb{E}\left[\operatorname{Tr}\left(\sum_{i,j=1}^{n} J^M(V_i)J^M(W_j)^T C_{ij}\right)\right] \\
&\quad + \mathbb{E}\left[\operatorname{Tr}\left(\sum_{i,j=1}^{n} (J(V_i) - MJ^M(V_i))MJ^M(W_j)^T C_{ij}\right)\right] \\
&\quad + \mathbb{E}\left[\operatorname{Tr}\left(\sum_{i,j=1}^{n} J(V_i)(J(W_j) - MJ^M(W_j))^T C_{ij}\right)\right]
\end{aligned}$$

Therefore it suffices to show that

$$\mathbb{E}\left[\operatorname{Tr}\left(\sum_{i,j=1}^{n} (J(V_i) - MJ^M(V_i))MJ^M(W_j)^T C_{ij}\right)\right] \leq \sqrt{\gamma(M,d,r)}\,\omega_{SDP}$$

and

$$\mathbb{E}\left[\operatorname{Tr}\left(\sum_{i,j=1}^{n} J(V_i)(J(W_j) - MJ^M(W_j))^T C_{ij}\right)\right] \leq \sqrt{\gamma(M,d,r)}\,\omega_{SDP}.$$

For the second inequality, we see by condition (1) of Definition 2.1 that

$$\mathbb{E}\left[J(V_i)J(V_i)^T\right] = V_i V_i^T = I_{d\times d} \preceq 1 I_{d\times d},$$

and by Lemma 2.3 we have that

$$\mathbb{E}\left[\left(J(W_j) - MJ^M(W_j)\right)\left(J(W_j) - MJ^M(W_j)\right)^T\right] \leq \gamma(M,d,r) I_{d\times d}.$$

The desired inequality then follows by Lemma 2.4.

The first inequality follows in a similar fashion, upon noting that $\left\|MJ^M(V)\right\|_\infty \leq \|J(V)\|_\infty$.

□

3. Gaussian Rounding

In this section we consider the particular rounding procedure

$$J(V) = VR$$

where $R \in \mathbb{R}^{2nd \times r}$ has i.i.d $\mathcal{N}(0, \frac{1}{r})$ entries.

Lemma 3.1. *J is an orthogonal rounding procedure.*

Proof. We immediately see that

$$\mathbb{E}\left[J(V)J(W)^T\right] = \mathbb{E}\left[VRR^TW^T\right] = V\mathbb{E}\left[RR^T\right]W^T = VI_{2nd\times 2nd}W^T = VW^T$$

so condition (1) in Definition 2.1 is satisfied.

Also J is computable in polynomial time in the size of V, so condition (3) is satisfied.

For condition (2), we will provide an explicit bound for $\gamma(M, d, r)$.

First let $V^T = QT$ be the QR-decomposition of V^T, where $Q \in \mathcal{O}(2nd)$ and $T \in \mathbb{R}^{2nd\times d}$ is upper triangular. Letting T_1 be the top $d \times d$ block of T, we note that all other entries of T are 0. Then since T_1 is upper triangular and $T_1^T T_1 = T^T T = T^T Q^T QT = VV^T = I_{d\times d}$, we see that T_1 is in fact diagonal with all diagonal entries equal to ± 1. We may further assume that $T_1 = I_{d\times d}$ by modifying Q to correct all the -1's.

By the rotational invariance of the Gaussian, it follows that

$$J(V) = VR = VQQ^TR = T^TQ^TR \sim T^TR = R_1$$

where R_1 is the first $d \times r$ block of R, so

$$\gamma(M, d, r) = 2\int_M^\infty \mathbb{P}\left\{\|R_1\|_\infty \geq t\right\}(t - M)dt$$

Now from Ref. [6] we have that

$$\mathbb{P}\left\{\|J(V)\|_\infty \geq t\right\} = \mathbb{P}\left\{\|R_1\|_\infty \geq t\right\} \leq 2e^{-\frac{(\sqrt{r}(t-1)-\sqrt{d})^2}{2}}$$

for all $t \geq 1 + \sqrt{\frac{d}{r}}$.

Therefore for $M \geq 1 + \sqrt{\frac{d}{r}}$,

$$\gamma(M, d, r) \leq 4\int_M^\infty e^{-\frac{(\sqrt{r}(t-1)-\sqrt{d})^2}{2}}(t - M)dt = \frac{2}{r}\left(2e^{-\frac{\phi^2}{2}} - \sqrt{2\pi}\phi\left(1 - \operatorname{erf}\left(\frac{\phi}{\sqrt{2}}\right)\right)\right)$$

where $\phi = (M - 1)\sqrt{r} - \sqrt{d}$. Therefore $\gamma(M, d, r) \to 0$ as $M \to \infty$, so condition (2) is satisfied. $\square$

By Theorem 2.5, we conclude that this rounding procedure provides an approximation ratio of

$$\beta(d, r) = \max_{M>0} \frac{1 - 2\sqrt{2\int_M^\infty \mathbb{P}\left\{\|S\|_\infty \geq t\right\}(t - M)dt}}{M^2} = \max_{M>0} \frac{1 - 2\sqrt{\mathbb{E}\left[\left((\|S\|_\infty - M)_+\right)^2\right]}}{M^2}$$

where $S \in \mathbb{R}^{d\times r}$ has i.i.d $\mathcal{N}(0, \frac{1}{r})$ entries.

From the discussion in Lemma 3.1, we obtain a lower bound of

$$\beta(d,r) \geq \max_{M \geq 1+\sqrt{\frac{d}{r}}} \frac{1 - \frac{4}{\sqrt{r}}\sqrt{e^{-\frac{\phi^2}{2}} - \sqrt{\frac{\pi}{2}}\phi\left(1 - \operatorname{erf}\left(\frac{\phi}{\sqrt{2}}\right)\right)}}{M^2} \tag{*}$$

where $\phi = (M-1)\sqrt{r} - \sqrt{d}$.

For any fixed d this bound converges to 1 as $r \to \infty$, and for $d = r \to \infty$ this bound converges to $\frac{1}{4}$. This immediately shows that for any fixed d, $\beta(d,r) \geq \frac{1}{2\sqrt{2}}$ for all but finitely many r (although we suspect that $\beta(d,d)$, which is the ratio for the bipartite problem over the orthogonal group, never exceeds $\frac{1}{2\sqrt{2}}$).

In fact we do not need this lower bound when $d = 1$, because in this case it is easy to explicitly compute $\beta(1,r)$. Indeed when $d = 1$, $\sqrt{r}\,\|S\|_\infty$ has the distribution $\chi(r)$, so

$$\mathbb{P}\left\{\|S\|_\infty \geq t\right\} = 1 - P\left(\frac{r}{2}, \frac{rt^2}{2}\right)$$

where $P(s,x) = \dfrac{\gamma(s,x)}{\Gamma(s)} = \dfrac{\int_0^x u^{s-1}e^{-u}du}{\int_0^\infty u^{s-1}e^{-u}du}$ is the regularized Gamma function.

Therefore

$$\beta(1,r) = \max_{M>0} \frac{1 - 2\sqrt{2\int_M^\infty \left(1 - P\left(\frac{r}{2}, \frac{rt^2}{2}\right)\right)(t-M)dt}}{M^2} \tag{†}$$

Using (†) when $d = 1$ and (*) when $d > 1$, we obtain the following table:

d/r	1	2	3	∞
1	0.267379	0.407071	0.495159	1
2	-	0.056329	0.078930	1
3	-	-	0.070617	1
∞	-	-	-	0.25

The values for $(d,r) = (2,2), (2,3)$, and $(3,3)$ are lower bounds for $\beta(d,r)$, and it appears that they are quite loose. By running a Monte-Carlo simulation on the expression

$$\beta(d,r) = \max_{M>0} \frac{1 - 2\sqrt{\mathbb{E}\left[\left((\|S\|_\infty - M)_+\right)^2\right]}}{M^2}$$

we recover the following approximate values:

d/r	1	2	3
1	0.2674	0.4071	0.4952
2	-	0.2419	0.3092
3	-	-	-

4. Future Work

In the future we plan to compute more values of $\beta(d, r)$ both by developing explicit expressions and through simulation, to explore the potential monotonicity of $\beta(d, r)$ along both dimensions (the diagonal does not appear to be monotonic), to test the rounding algorithm in practice, to develop integrality gaps for the problem, to consider other orthogonal rounding procedures, and to consider the bipartite problem over the unitary group / complex Stiefel manifold.

We also believe that there exists a 1/log(n) approximation ratio for the general (non-bipartite, non-little) Grothendieck problem, and we hope to show this as well as that a constant factor ratio is impossible to achieve in polynomial time (modulo computational hardness assumptions).

References

[1] N. Alon and A. Naor. "Approximating the Cut-Norm via Grothendieck's Inequality", *Proceedings of the* 36^{th} *Annual ACM Symposium on Theory of Computing (STOC)*, 2004.

[2] A. Bandeira, C. Kennedy, and A. Singer. "Approximating the Little Grothendieck Problem over the Orthogonal and Unitary Groups." *In review*, 2013.

[3] M.X. Goemans and D.P. Williamson. "Improved Approximation Algorithms for Maximum Cut and Satisfiability Problems Using Semidefinite Programming." *Journal of the Association for Computing Machinery,* 42:1115-1145, 1995.

[4] A. Naor, O. Regev, and T. Vidick. "Efficient Rounding for the Noncommutative Grothendieck Inequality." *Proceedings of the* 45^{th} *Annual ACM Symposium on Theory of Computing (STOC)*, 2013.

[5] L. Vanderberghe and S. Boyd. "Semidefinite Programming." *SIAM Review,* 38:49-95, 1996.

[6] R. Vershynin. "Introduction to the Non-Asymptotic Analysis of Random Matrices." *Ch. 5 of Compressed Sensing, Theory, and Applications*, Cambridge University Press, 210-268, 2012.

Department of Mathematics, Princeton University, Princeton, NJ 08544

E-mail address: airiza@princeton.edu