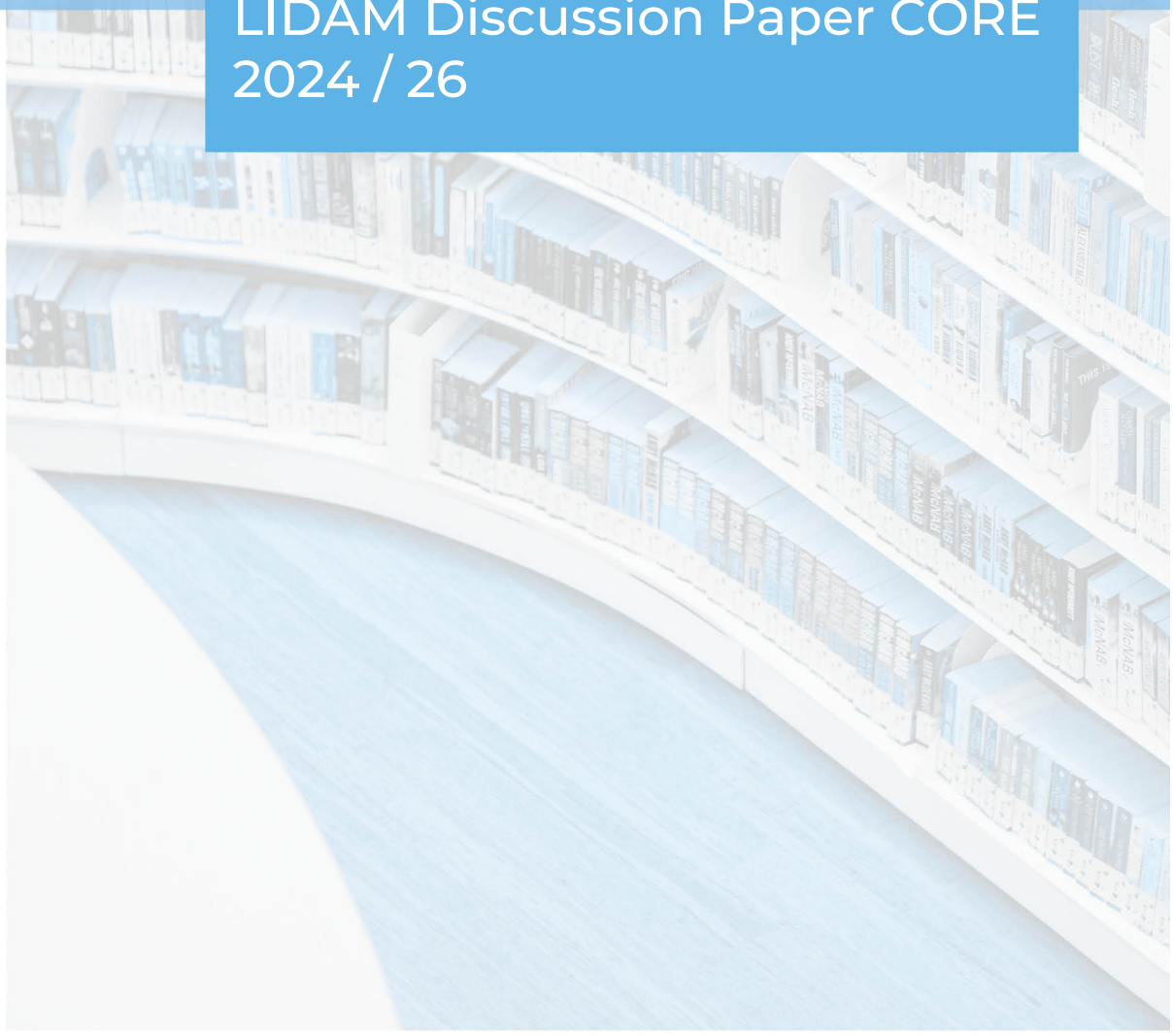
# CONVEX QUARTIC PROBLEMS: HOMOGENIZED GRADIENT METHOD AND PRECONDITIONING

Radu-Alexandru Dragomir, Yurii Nesterov

LIDAM Discussion Paper CORE
2024 / 26



UCLouvain

CORE | LIDAM
Louvain Institute of Data Analysis and Modeling in economics and statistics

**CORE**

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: immaq-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/core/core-discussion-papers.html

# Convex quartic problems: homogenized gradient method and preconditioning

Radu-Alexandru Dragomir[1] and Yurii Nesterov[2]

[1]EPFL, Switzerland
[2]CORE, UCLouvain, Belgium

April 24, 2024

**Abstract**

We consider a convex minimization problem for which the objective is the sum of a homogeneous polynomial of degree four and a linear term. Such task arises as a subproblem in algorithms for quadratic inverse problems with a difference-of-convex structure. We design a first-order method called *Homogenized Gradient*, along with an accelerated version, which enjoy fast convergence rates of respectively $\mathcal{O}(\kappa^2/K^2)$ and $\mathcal{O}(\kappa^2/K^4)$ in relative accuracy, where $K$ is the iteration counter. The constant $\kappa$ is the quartic condition number of the problem.

Then, we show that for a certain class of problems, it is possible to compute a preconditioner for which this condition number is $\sqrt{n}$, where $n$ is the problem dimension. To establish this, we study the more general problem of finding the best quadratic approximation of an $\ell_p$ norm composed with a quadratic map. Our construction involves a generalization of the so-called Lewis weights.

## 1 Introduction

### 1.1 Problem setup

We consider the convex minimization problem

$$\min_{x \in \mathcal{K}} f(x) \triangleq \rho(x) - \langle c, x \rangle \tag{P}$$

where $\mathcal{K}$ is a closed convex cone of a Euclidean space $E$, $c$ is a nonzero vector of $E$, and $\rho$ is a convex quartic form, meaning that $\rho$ is convex and

$$\rho(x) = Q[x, x, x, x],$$

for some symmetric 4-linear form $Q : E^4 \to \mathbb{R}$. We assume that $\rho$ satisfies the following bounds with respect to a given Euclidean norm $\|\cdot\|$:

$$\alpha^2 \|x\|^4 \leq \rho(x) \leq \beta^2 \|x\|^4, \quad \forall x \in E, \tag{1.1}$$

where $0 < \alpha \leq \beta$. We call $\kappa \triangleq \frac{\beta}{\alpha}$ the *quartic condition number* of $\rho$. Problem (P) arises as a subproblem in some algorithms for solving a large class of quartic optimization problems, as we detail in Section 2. Applications include quadratic least squares, phase retrieval, distance matrix completion and matrix factorization.

We propose to study first-order methods for solving (P). In standard gradient methods, the objective gap is bounded as $\mathcal{O}(LR^2/K)$, where $K$ is the iteration counter, $R$ is the initial distance to the optimal point and $L$ is the Lipschitz constant of the gradient; accelerated methods allow the convergence speed to be improved to $\mathcal{O}(LR^2/K^2)$ (Nesterov, 1983). However, these methods require the gradient of $f$ to be globally Lipschitz continuous, which is not the case for quartic function $\rho$. This can be remedied by constraining the optimization on a bounded set, but this generally leads to a highly over-estimated Lipschitz constant.

In this work, we adopt a different approach by exploiting the polynomial quartic structure of convex function $\rho$ and the regularity condition (1.1).

### 1.2 Contributions and outline

Our main goal is to design first-order methods which enjoy fast convergence rates for solving Problem (P). We list below our main contributions.

- **Regularity properties of convex quartic functions** (Section 3): using the polynomial structure of $\rho$ and the condition (1.1), we establish two crucial properties for our analysis. First, we show that $\rho$ satisfies a *uniform convexity property* of degree 4. Then, we prove that the function $\sqrt{\rho}$ is convex and has Lipschitz continous gradient.

- **Homogenized gradient descent** (Section 4): as a next step, we show that a solution to Problem (P) can be obtained by solving a *homogenized problem*

$$\min_{\substack{y \in \mathcal{K} \\ \langle c, y \rangle = 1}} \sqrt{\rho(y)}. \tag{H}$$

  As $\sqrt{\rho}$ is convex and has Lipschitz gradient, we can apply projected gradient descent to (H). The uniform convexity of $\rho$ allows to establish faster convergence rates than in the standard smooth convex case. We prove that homogenized gradient descent enjoys a $\mathcal{O}(\kappa^2/K^2)$ rate, and that an accelerated technique allows to improve this speed to $\mathcal{O}(\kappa^2/K^4)$. The key to obtain this fast sublinear rate is the *restart* mechanism, which has been studied before in situations similar to uniform convexity (Nemirovski and Yudin, 1983; Roulet and d'Aspremont, 2017; Nesterov, 2020).

- **Optimal preconditioning** (Section 5): as the efficiency of our methods depend on the condition number $\kappa$, we look to improve it by choosing an appropriate Euclidean norm of the form $\|\cdot\|_B$ induced by a positive definite operator $B$. We study the case where $\rho$ is of the form

$$\rho(x) = \sum_{i=1}^{m} \langle x, B_i x \rangle^2. \tag{1.2}$$

  This structure arises in our main application of interest, the *quadratic inverse problems*. We show that there exists a Euclidean norm $\|\cdot\|_{B^*}$ induced by an operator $B^*$ for which $\kappa = \sqrt{n}$. Our analysis can be applied to the problem of approximating the function $x \mapsto (\sum_{i=1}^m \langle B_i x, x \rangle^p)^{1/p}$, $p \geq 1$, by a Euclidean norm. We show that there exists a norm with condition number $n^{1-1/p}$, which is not improvable in general. This norm is induced by an operator $B^* = \sum_{i=1}^m \tau_i^* B_i$, where $\tau^*$ satisfies a given fixed-point equation.

  The coefficients $\tau_i$ can be seen as a generalization of the called *Lewis weights* used in $\ell_p$ subspace approximation Lewis (1978).

We then show that $\tau^*$ can be computed with a fixed-point algorithm, which runs in $\tilde{\mathcal{O}}(mrn^2)$ time (the notation $\tilde{\mathcal{O}}$ hiding logarithmic factors), where $n = \dim E$ and $r$ is the maximal rank of the $B_i$'s. Using $B^*$ allows significant gains over the simpler choice $\overline{B} = \sum_{i=1}^m B_i$ if $m \gg n$ and the set of matrices $B_1, \dots B_m$ has *high coherence*.

- **Numerical experiments** (Section 6): finally, we apply our algorithms to several problems of the form (1.2). We generate instances of $\rho$ with various condition number and coherence values in order to demonstrate the impact of homogenization, acceleration and preconditioning.

## 1.3 Related work

### 1.3.1 Polynomial growth and fast convergence rates

The use of a lower approximation of the objective function in order to improve convergence rates has been long studied in first-order optimization. When the lower bound is quadratic, this amounts to strong convexity and allows to obtain linear convergence. If it is polynomial of some other degree, linear convergence is out of reach but faster sublinear rates can be obtained. Several variants of polynomial lower bound have been studied, such as *uniform convexity* (Iouditski and Nesterov, 2014; Nesterov, 2020), Kurdyka–Łojasiewicz property (Bolte et al., 2014, 2017; Roulet and d'Aspremont, 2017) or other conditions (Freund and Lu, 2018); see also (Doikov, 2022) for lower complexity bounds in this setting. Compared to previous work, the novelty of our setting is that the uniform convexity property does not hold for the objective function that we are trying to minimize, but its square (see Section 4.2). Minimization of convex quartic polynomials has also been tackled by higher-order tensor methods (Bullins, 2018; Nesterov, 2020).

### 1.3.2 Bregman methods

Another approach which has been applied to quartic problems is relatively-smooth minimization (Bauschke et al., 2017; Lu et al., 2018). This method relies on the choice of a reference function $h$ whose Hessian upper bounds that of $f$, allowing to obtain possibly a better efficiency than the standard Euclidean gradient method. In the case of quartic functions, this technique has been applied by choosing $h$ to be a simple polynomial of degree 4 (Bolte et al., 2018; Mukkamala and Ochs, 2019; Dragomir et al., 2021a). For convex functions, such scheme enjoys a convergence rate of $\mathcal{O}(LD/K)$, where $L$ is the relative Lipschitz constant and $D$ the initial Bregman distance to the optimum.

A main drawback of Bregman methods is that they are difficult to accelerate: the lower bound by (Dragomir et al., 2021b) guarantees that no Bregman-type algorithm can do better than $\mathcal{O}(LD/K)$ for generic reference functions $h$. Even with additional regularity assumptions, known acceleration results are limited to local or asymptotic regimes (Hanzely et al., 2021; Hendrikx et al., 2020). Another caveat is that relatively-smooth minimization does not allow to use the lower bound on $\rho$ in (1.1) in order to improve convergence guarantees[1].

### 1.3.3 Preconditioning and optimal quadratic norms

The idea of finding the best quadratic norm to approximate a convex set can be traced back to John (John, 1948). This amounts to computing the *minimum-volume ellipsoid* enclosing a given

[1] One could wonder if the term $\alpha\|x\|^4$ implies a *relative strong convexity* (Lu et al., 2018) with respect to a well-chosen quartic function $h$. In turns out that this is not true in general, apart from trivial choices such as $h = \rho$. This motivates the need to consider the weaker notion of *uniform convexity*, which will not be enough to guarantee linear convergence but will lead to improved sublinear rates.

set. Efficient procedures for computing such ellipsoid have been proposed in (Khachiyan, 1996; Anstreicher, 1999); see also (Todd, 2016) for a recent survey. This principle has been applied to linear programming (Nesterov, 2008). In our case, the direct applications of such methods would yield a quartic condition number of $n$, where $n$ is the dimension of $E$. In our work, we propose a variant specialized to quartic functions, allowing to compute a quadratic norm for which the condition number of $\rho$ is reduced to $\sqrt{n}$.

#### 1.3.4 Lewis weights

In the case where the matrices $B_i$ are of rank 1, i.e. $B_i = a_i a_i^T$ for some $a_i \in \mathbb{R}^n$, then the weights $\tau^*$ satisfying (**??**) are called the *Lewis weights* of the matrix $A = [a_1^T \dots a_m^T]^T \in \mathbb{R}^{m \times n}$. They were initially introduced in (Lewis, 1978) and studied further in (Bourgain et al., 1989; Talagrand, 1995; Cohen and Peng, 2015). These weights are used as sampling probabilities to approximate the $\ell_p$-norm $\|Ax\|_p$ by selecting a random subset of the rows of $A$. In our work, we show that the same quantities can be used to build an approximation of functions of the type $\|Ax\|_{2p}$ by a Euclidean norm $\|x\|_B = \langle Bx, x\rangle^{\frac{1}{2}}$. The algorithm that we use for computing $\tau^*$ is inspired by a fixed point scheme in (Cohen and Peng, 2015).

### 1.4 Notation and definitions

Throughout the paper, $\mathcal{K}$ denotes a convex cone of a Euclidean space $E$. We assume that the norm $\|\cdot\|$ is induced by a self-adjoint positive definite operator $B : E \to E^*$ as

$$\|x\| = \|x\|_B = \langle Bx, x\rangle,$$

for any $x \in E$, where $\langle \cdot, \cdot \rangle$ is the canonical inner product. We will write $\|\cdot\|_B$ when we want to insist on a specific choice of matrix $B$, and $\|\cdot\|$ when it is clear from context.

A function $Q : E^4 \to \mathbb{R}$ is a 4-symmetric linear map if it is linear in each variable and invariant with respect to permutation of the variables. We use the short notation $Q[x]^4 = Q[x, x, x, x]$ and $Q[x]^2[y]^2 = Q[x, x, y, y]$.

The notation $[t]_+ = \max(t, 0)$ denotes the positive part of a real number $t$. We say that a differentiable function $g : E \to \mathbb{R}$ is *uniformly convex* of degree $p \geq 2$ with parameter $\sigma$ if for any $x, y \in E$ we have

$$g(x) - g(y) - \langle \nabla g(y), x - y\rangle \geq \frac{\sigma}{p}\|x - y\|^p.$$

## 2 Motivation and examples

In this section, we give examples of relevant applications involving minimization of problems of the form of (P).

### 2.1 DC algorithm for quadratic inverse problems

The auxiliary problems of the form (P) arise in the algorithms for *quadratic inverse problems*

$$\min_{x \in \mathcal{K}} F(x) \triangleq \sum_{i=1}^{m} \left(\langle x, B_i x\rangle - d_i\right)^2 \tag{QIP}$$

where $B_1 \dots B_m$ are symmetric linear operators on $E$ and $d_1 \dots d_m \in \mathbb{R}$ some target values. Problem (QIP) is generally nonconvex and occurs in various tasks of data science and signal processing; we

detail some examples below.

Expanding the square, $F$ rewrites as a difference of two functions $\rho, \phi$:

$$F(x) = \underbrace{\sum_{i=1}^m \langle x, B_i x\rangle^2}_{\rho(x)} - \underbrace{\sum_{i=1}^m \left(2d_i\langle x, B_i x\rangle - d_i^2\right)}_{\phi(x)}.$$

We make the following assumption:

$$\rho \text{ and } \phi \text{ are } \textbf{convex functions}.$$

This holds for instance if $B_1, \dots B_m$ are positive semidefinite operators and $d_i \geq 0$ for $i = 1 \dots m$. We say that $F$ has a **difference-of-convex** (DC) structure. Convexity of $\phi$ allows to design a convex global majorant of $F$ around any point $\overline{x}$ with the inequality

$$F(x) \leq \rho(x) - \phi(\overline{x}) - \langle \nabla\phi(\overline{x}), x - \overline{x}\rangle.$$

Then the DC algorithm (Tao and Souad, 1986; An and Tao, 2005) simply consists in successive minimization of this global majorant: choose $x_0 \in \mathcal{K}$ and compute for $k = 0, 1, \dots$

$$x_{k+1} \in \operatorname*{argmin}_{x\in\mathcal{K}} \rho(x) - \langle \nabla\phi(x_k), x\rangle. \tag{DC}$$

We recognize here a problem of the form (P), with $c = \nabla\phi(x_k)$ and $\rho$ is the convex quartic polynomial

$$\rho(x) = \sum_{i=1}^m \langle x, B_i x\rangle^2. \tag{2.1}$$

The corresponding symmetric 4-linear form is $Q[u, v, w, z] = \sum_{i=1}^m \langle u, B_i v\rangle\langle w, B_i z\rangle$.

The DC algorithm has mostly been studied in the context of polyhedral functions. It is quite general as any function with Lipschitz continuous gradient can be proven to have a DC decomposition. Its efficiency depends on how well we can solve the subproblems defining $x_{k+1}$ in (DC). In this work, we show that the case where $\rho$ is a convex quartic polynomial is a favorable situation.

We show that, when the operators $B_i$ are positive semidefinite, function $\rho$ satisfies the quartic conditioning assumption (1.1). In particular, the constant $\alpha$ is positive when the operator $\sum_{i=1}^m B_i$ is positive definite.

**Proposition 2.1.** *Assume that the operators $B_1, \dots B_m$ are all positive semidefinite and denote $\bar{B} = \sum_{i=1}^m B_i$. Then the function $\rho$ defined in (2.1) satisfies*

$$\frac{\lambda_{\min}(\bar{B})^2}{m}\|x\|_2^4 \leq \rho(x) \leq \lambda_{\max}(\bar{B})^2\|x\|_2^4,$$

*for $x \in E$, where $\lambda_{\min}$ and $\lambda_{\max}$ denote the smallest and largest eigenvalues.*

*Proof.* The lower bound follows from Cauchy-Schwarz:

$$\left(\sum_{i=1}^m \langle x, B_i x\rangle\right)^2 \leq m\sum_{i=1}^m \langle x, B_i x\rangle^2.$$

For the upper bound, since $\langle B_i x, x\rangle \geq 0$ for each $i$ we have

$$\sum_{i=1}^m \langle x, B_i x\rangle^2 \leq \left(\sum_{i=1}^m \langle x, B_i x\rangle\right)^2.$$

We conclude by using $\lambda_{\min}(\bar{B})\|x\|_2^2 \leq \sum_{i=1}^m \langle x, B_i x\rangle \leq \lambda_{\max}(\bar{B})\|x\|_2^2$.

□

Therefore, the quartic condition number of $\rho$ is upper bounded by $\sqrt{m}\kappa_2(\bar{B})$, where $\kappa_2(\bar{B})$ is the condition number in the usual sense of operator $\bar{B}$. In Section 5, we refine this estimate using the *coherence* of the set $(B_1, \dots B_m)$.

We now list a few examples.

**Example 1: phase retrieval** We wish to recover a complex signal $x^* \in \mathbb{C}^N$ from phaseless quadratic measurements $d_i = |\langle q_i, x^*\rangle|^2$ where $q_i \in \mathbb{C}^N$ for $i = 1 \dots m$. This is a fundamental signal processing problem with applications in optimal imaging, cristallography and astronomy (Shechtman et al., 2014; Candès et al., 2015; Sun et al., 2017). A popular method for performing phase retrieval is to solve the nonconvex optimization problem

$$\min_{x\in\mathbb{C}^N} \frac{1}{m}\sum_{i=1}^m (|\langle q_i, x\rangle|^2 - d_i)^2. \tag{2.2}$$

The DC structure arises as $\phi(x) = \frac{1}{m}\sum_{i=1}^m (2d_i|\langle q_i, x\rangle|^2 - d_i^2)$ is a convex quadratic and the quartic part is

$$\rho(x) = \frac{1}{m}\sum_{i=1}^m |\langle q_i, x\rangle|^4.$$

The operator $B_i$ is $q_i q_i^H$ (where $q^H$ denotes Hermitian transpose). Since the vectors $q_1, \dots q_m$ are typically obtained from overcomplete Fourier bases or Gaussian distributions, the sum $\sum_{i=1}^m B_i$ is generally invertible and well-conditioned (Candès et al., 2015).

**Example 2: distance matrix completion (DMC)** In this problem, we wish to recover the position of $N$ points $x_1^*, \dots, x_N^* \in \mathbb{R}^r$ from the partial knowledge of the pairwise distances $d_{ij} = \|x_i^* - x_j^*\|_2^2$ for $i, j \in \Omega$, where $\Omega \subset \{1 \dots N\} \times \{1 \dots N\}$. EDMC has applications in sensor network localization and biology; see Fang and O'Leary (2012); Qi and Yuan (2013); Dokmanic et al. (2015) and references therein. There are several ways to solve this task, including semidefinite programming; however, for large-scale problems, the following nonconvex formulation is popular due to its smaller dimension:

$$\min_{\substack{X\in\mathbb{R}^{N\times r}\\ X^T\mathbf{1}=0}} \sum_{(i,j)\in\Omega} \left(\|x_i - x_j\|_2^2 - d_{ij}\right)^2, \tag{DMC}$$

where $x_1, \dots x_N$ are the rows of matrix $X$, and the centering constraint allows to remove the translation invariance. The convex quartic function $\rho$ can be written in this case

$$\rho(X) = \sum_{(i,j)\in\Omega} \|x_i - x_j\|_2^4 + \frac{1}{N^2}\Big\|\sum_{i=1}^N x_i\Big\|^4.$$

The addition of the second term, which is 0 on the feasible set of (DMC), allows to make $\rho$ positive definite. The linear operator $\bar{B}$ as defined in Proposition 2.1 satisfies

$$\langle \bar{B}X, X\rangle = \sum_{(i,j\in\Omega)} \|x_i - x_j\|_2^2 + \frac{1}{N}\Big\|\sum_{i=1}^N x_i\Big\|^2$$

$$= \sum_{k=1}^{r} \langle \left( \mathcal{L}(\Omega) + \frac{1}{N} \mathbf{1}_N \mathbf{1}_N^T \right) x_{\cdot k}, x_{\cdot k} \rangle,$$

where $x_{\cdot k}$ is the $k$-th column of $X$ and $\mathcal{L}(\Omega) \in \mathbb{R}^{N \times N}$ is the Laplacian matrix of the graph induced by $\Omega$ Chung (1997).

The smallest eigenvalue of $\mathcal{L}(\Omega)$ is 0 with eigenvector $\mathbf{1}_N$. The second smallest eigenvalue, which we denote $\lambda_1(\Omega)$, is often called the **graph spectral gap**. The largest eigenvalue is bounded by $4\bar{d}(\Omega)$, where $\bar{d}(\Omega)$ is the maximal degree of the graph. We deduce that the eigenvalues of operator $\bar{B}$ are located in $[\lambda_1(\Omega), 4\bar{d}(\Omega) + 1]$. It follows from Proposition 2.1 that for every $X \in \mathbb{R}^{N \times r}$,

$$\left( \frac{\lambda_1(\Omega)^2}{|\Omega| + 1} \right) \|X\|_F^4 \leq \rho(X) \leq (4\bar{d}(\Omega) + 1)^2 \|X\|_F^4.$$

The spectral gap $\lambda_1(\Omega)$ is nonzero if the graph is connected Chung (1997), which is a necessary condition for Problem (DMC) to be solvable.

**Example 3: symmetric nonnegative matrix factorization (SymNMF).** Let $M \in \mathbb{R}_+^{n \times n}$ be a nonnegative matrix, which we assume to be positive semidefinite. We wish to find a symmetric low-rank decomposition $M \approx XX^T$ where the factor $X \in \mathbb{R}_+^{n \times r}$ has nonnegative entries and $r$ is a small target rank. This task has relevant applications in graph clustering (He et al., 2011; Kuang et al., 2015). SymNMF is usually performed by solving the nonconvex problem

$$\min_{X \in \mathbb{R}_+^{n \times r}} \|XX^T - M\|_F^2 = \|XX^T\|_F^2 - \left( 2\langle MX, X \rangle - \|M\|_F^2 \right)$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Then $\mathcal{K} = \mathbb{R}_+^{n \times r}$ is the nonnegative orthant, the function $\phi(X) = 2\langle MX, X \rangle - \|M\|_F^2$ is a convex quadratic (since $M$ is positive semidefinite) and the quartic part is the convex function $\rho(X) = \|XX^T\|_F^2$. In this situation, Proposition 2.1 does not apply as the individual operators $B_i$ are not positive semidefinite. However, we can still estimate the quartic condition number. Using the properties of the Frobenius inner product we have

$$\rho(X) = \mathrm{Tr}(XX^T XX^T) = \mathrm{Tr}(X^T XX^T X) = \|X^T X\|_F^2 = \sum_{i,j=1}^{r} (X_{\cdot,i}^T X_{\cdot,j})^2,$$

where $X_{\cdot,i}$ is the $i$-th column of $X$. Hence $\rho$ satisfies the bounds

$$\rho(X) \leq \sum_{i,j=1}^{r} \|X_{\cdot,i}\|_F^2 \|X_{\cdot,j}\|_F^2 = \|X\|_F^4,$$

$$\rho(X) \geq \sum_{i=1}^{r} \|X_{\cdot,i}\|_F^4 \geq \frac{1}{r} \left( \sum_{i=1}^{r} \|X_{\cdot,i}\|_F^2 \right)^2 = \frac{1}{r} \|X\|_F^4.$$

The condition number is hence $\sqrt{r}$, which is small in most applications.

## 2.2 Statistical Bregman preconditioning for stochastic and distributed optimization

We consider the same setup as in the previous section, that is, a quadratic inverse problem of the form (QIP) with a difference-of-convex structure. We mention another technique which is especially suited for huge-scale and potentially distributed problems.

Assume that the measurement operators $B_1 \dots B_m$ are positive semidefinite and generated as

i.i.d. samples from a certain statistical distribution. Then, one can subsample this distribution by choosing a smaller index set $I \subset \{1 \dots m\}$ and form an approximation $\tilde{\rho}$ to $\rho$ as

$$\tilde{\rho}(x) = \frac{m}{|I|} \sum_{i \in I} \langle x, B_i x \rangle^2.$$

With the right statistical assumptions, one can prove that when $|I|$ is sufficiently high, *relative smoothness and strong convexity* hold with high probability:

$$\mu_m \nabla^2 \tilde{\rho}(x) \preceq \nabla^2 \rho(x) \preceq L_m \nabla^2 \tilde{\rho}(x), \tag{2.3}$$

where $\mu_m, L_m$ are the relative regularity constants. We refer the reader to (Shamir et al., 2014; Hendrikx et al., 2020) for detailed results on statistical preconditioning. Condition (2.3) allows to apply the Bregman (stochastic) gradient method with reference function $\tilde{\rho}$:

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathcal{K}} \langle g_k, x - x_k \rangle + \frac{1}{\lambda_k} \left( \tilde{\rho}(x) - \tilde{\rho}(x_k) - \langle \nabla \tilde{\rho}(x_k), x - x_k \rangle \right), \tag{2.4}$$

where $g_k$ is a stochastic unbiased estimate of $\nabla F(x_k)$, and $\lambda_k > 0$ is the step size.

The choice of $I$ involves a tradeoff: if $I = \{1 \dots m\}$ the problem (2.4) can be as hard to solve as the original one (QIP). However, choosing $I$ as a smaller subset can result in a significant performance gain compared to a standard Euclidean method. This approach can be advantageous in the context of distributed optimization, where the computation (2.4) is performed by a centralized server and the goal is to reduce the overall number of stochastic gradients computed as they involve costly communication among machines (Hendrikx et al., 2020).

Problem (2.4) is of the form (P) with $c = \nabla \tilde{\rho}(x_k) - \lambda_k g_k$, which is another motivation for our work.

# 3 Properties of convex quartic polynomials

Before studying methods for solving Problem (P), we need to establish some key preliminary results on convex quartic polynomials. Recall that we assume that $\rho$ induced by the symmetric 4-linear form $Q : E^4 \to \mathbb{R}$ as $\rho(x) = Q[x]^4$, and that $\rho$ is convex.

## 3.1 Convexity and smoothness of $\sqrt{\rho}$

We first recall the following fact about convex quartic forms.

**Lemma 3.1** (Theorem 1 in (Nesterov, 2022))**.** *Let $Q : E^4 \to \mathbb{R}$ be a symmetric 4-linear form for which the function $x \mapsto Q[x]^4$ is convex. Then for every $x, y \in E$ we have*

$$0 \leq Q[x]^2[y]^2,$$
$$\left(Q[x]^2[y]^2\right)^2 \leq Q[x]^4\, Q[y]^4, \tag{3.1}$$
$$\left(Q[x]^3[y]\right)^2 \leq Q[x]^4 Q[x]^2[y]^2. \tag{3.2}$$

This inequality implies that $\sqrt{\rho}$ has favorable properties for gradient methods.

**Proposition 3.2.** *The function $g = \sqrt{\rho}$ is convex, differentiable and has Lipschitz continuous gradients with constant $6\beta$.*

*Proof.* For $x \neq 0$, $\rho(x) > 0$ by (1.1) and therefore $g$ is twice differentiable at $x$. Its derivative along a direction $h \in E$ is

$$\langle \nabla g(x), h \rangle = \frac{2Q[x]^3[h]}{(Q[x]^4)^{1/2}}. \tag{3.3}$$

Its second derivative writes for $h \in E$

$$g''(x)[h,h] = \frac{6\,Q[x]^2[h]^2}{(Q[x]^4)^{1/2}} - \frac{4\left(Q[x]^3[h]\right)^2}{(Q[x]^4)^{3/2}}.$$

Equation (3.2) implies that $g''(x)[h,h] \geq \frac{2\,Q[x]^2[h]^2}{(Q[x]^4)^{1/2}} \geq 0$, which proves convexity on $E \setminus \{0\}$. Since $g$ is continuous at 0, convexity on $E$ follows.

Let us now prove Lipschitz continuity of gradients. Using Equation (3.1) and the bound (1.1) yields that for $x \neq 0$,

$$g''(x)[h,h] \leq \frac{6\,Q[x]^2[h]^2}{(Q[x]^4)^{1/2}} \leq 6\left(Q[h]^4\right)^{1/2} \leq 6\beta\|h\|^2.$$

This proves that $\|\nabla g(x) - \nabla g(y)\| \leq 6\beta\|x-y\|$ for every $x, y \in E$ such that 0 does not belong to the segment $[x,y]$.

Consider now the situation at 0. Since $|g(0+h)| = |\sqrt{\rho(h)}| \leq \beta\|h\|^2$, $g$ is differentiable at 0 and $\nabla g(0) = 0$. Let us prove that $\nabla g$ is continuous at 0. Using the expression (3.3) and Lemma 3.1, we have for $x \neq 0$ and $h \in E$

$$|\langle \nabla g(x), h \rangle| \leq 2\left(Q[x]^4 Q[h]^4\right)^{1/4} \leq 2\beta\|x\|\|h\|$$

and therefore $\|\nabla g(x)\| \leq 2\beta\|x\|$, implying that $\nabla g$ is continuous at 0. Hence, by continuity and the previous point we have $\|\nabla g(x) - \nabla g(0)\| \leq 6\beta\|x\|$ for every $x \in E$. It remains to deal with the situation where $0 \in [x,y]$. In this case we have

$$\|\nabla g(x) - \nabla g(y)\| \leq \|\nabla g(x)\| + \|\nabla g(y)\| \leq 6\beta(\|x\| + \|y\|) = 6\beta\|x-y\|.$$

□

### 3.2 Uniform convexity of $\rho$

We show that $\rho$ satisfies the *uniform convexity* property of degree 4.

**Proposition 3.3.** *The function $\rho$ is uniformly convex of degree 4 with constant $4\alpha^2/3$: for every $x, y \in E$,*

$$\rho(x) - \rho(y) - \langle \rho'(y), x-y \rangle \geq \frac{\alpha^2}{3}\|x-y\|^4.$$

*Proof.* Let $h = x - y$. By expanding $Q[y+h]^4$ we get

$$\begin{aligned}
\rho(y+h) - \rho(y) - \langle \rho'(y), h \rangle &= Q[y+h]^4 - Q[y]^4 - 4Q[y]^3[h] \\
&= Q[h]^4 + 6Q[y]^2[h]^2 + 4Q[y][h]^3 \\
&\geq \frac{\alpha^2}{3}\|h\|^4 + \frac{2}{3}Q[h]^4 + 6Q[y]^2[h]^2 + 4Q[y][h]^3
\end{aligned}$$

We show that the residual is nonnegative using the inequality $a^2 + b^2 \geq 2ab$ and Lemma 3.1:

$$\frac{2}{3}Q[h]^4 + 6Q[y]^2[h]^2 + 4Q[y][h]^3 \geq 2\sqrt{4Q[h]^4Q[y]^2[h]^2} - |4Q[y][h]^3| \geq 0.$$

□

# 4 Homogenized gradient descent

We now design gradient methods for solving (P). Using the results from the previous sections, we first show that it can be transformed into a more favorable *homogenized* problem, which satisfies the classical assumptions for smooth convex minimization.

From now on, we make an assumption ensuring that problem (P) is not trivial.

**Assumption 4.1.** *There exists $x \in \mathcal{K}$ such that $\langle c, x\rangle > 0$.*

In other words, $c$ does not belong to the polar cone $\mathcal{K}^\circ$. If this was the case, the problem would minimized at 0.

## 4.1 Homogenization

Recall that $f(x) = \rho(x) - \langle c, x\rangle$. Let $f_{\text{hom}}$ the *homogenization* of $f$ be defined for $y \in E$ as

$$f_{\text{hom}}(y) = \min_{s\in\mathbb{R}_+} f(sy)$$

As $f(sy) = s^4\rho(y) - s\langle c, y\rangle$, the one-dimensional minimization in $s$ can be computed easily. Writing $s(y)$ the solution to this minization, we have for $y \neq 0$

$$\begin{aligned} s(y) &= \operatorname*{argmin}_{s\geq 0} f(sy) = \left(\frac{[\langle c, y\rangle]_+}{4\rho\,(y)}\right)^{1/3}, \\ f_{\text{hom}}(y) &= f(s(y)y) = \frac{-3}{4^{4/3}}\left(\frac{[\langle c, y\rangle]_+^4}{\rho(y)}\right)^{1/3}. \end{aligned} \tag{4.1}$$

Since Problem (P) is defined on a cone $\mathcal{K}$, we have

$$\min_{x\in\mathcal{K}} f(x) = \min_{y\in\mathcal{K}} \min_{s\geq 0} f(sy) = \min_{y\in\mathcal{K}} f_{\text{hom}}(y),$$

and the corresponding minimizers are

$$\operatorname*{argmin}_{x\in\mathcal{K}} f(x) = \left\{ s(y)y \,:\, y \in \left\{ \operatorname*{argmin}_{y\in\mathcal{K}} f_{\text{hom}}(y)\right\}\right\}. \tag{4.2}$$

Since $f_{\text{hom}}(y) = 0$ if $\langle c, y\rangle \leq 0$, we can restrict the minimization to the set $\{y \in \mathcal{K} : \langle c, y\rangle > 0\}$, which is nonempty by Assumption 4.1. Furthermore, note that $f_{\text{hom}}$ is a homogeneous function of degree 0. Since it is independent to rescaling, we can restrict to the set where $\langle c, y\rangle = 1$ and write

$$\min_{y\in\mathcal{K}} f_{\text{hom}}(y) = \min_{y\in\mathcal{K}} \frac{-3}{4^{4/3}}\left(\frac{\langle c, y\rangle^4}{\rho(y)}\right)^{1/3} = \min_{\substack{y\in\mathcal{K}\\ \langle c,y\rangle=1}} \frac{-3}{4^{4/3}\rho(y)^{1/3}}$$

Note that the minimizer satisfies $\rho(y^*) > 0$. Since any increasing transformation of the objective function preserves the minimizer, this problem has the same solution as

$$\min_{\substack{y\in\mathcal{K}\\ \langle c,y\rangle=1}} \sqrt{\rho(y)}. \tag{H}$$

We summarize this discussion in the following result, which relates the quality of a solution to (H) to that of the original problem (P) in *relative accuracy*.

**Proposition 4.2** (Homogenization). *Let $y^*$ be a minimizer of (H). Then*

*(i) $x^* = s(y^*)y^*$ is a minimizer of (P),*

*(ii) for any $y \in \mathcal{K}$ such that $\langle c, y\rangle = 1$, the point $x = s(y)y$ satisfies*

$$\frac{f(x) - f(x^*)}{|f(x^*)|} \leq \frac{2}{3}\frac{\sqrt{\rho(y)} - \sqrt{\rho(y^*)}}{\sqrt{\rho(y^*)}}$$

*Proof.* If $y^*$ is minimizer of (H), then it is also a minimizer of $f_{\text{hom}}$ on $\mathcal{K}$, and by (4.2) the point $x^* = s(y^*)y^*$ is a minimizer of (P). Then, for $y \in \mathcal{K}$ such that $\langle c, y\rangle = 1$ we have

$$\begin{aligned} f(s(y)y) - f(s(y^*)y^*) = \;& f_{\text{hom}}(y) - f_{\text{hom}}(y^*) = \tfrac{3}{4^{4/3}}\left(\tfrac{-1}{\rho(y)^{1/3}} - \tfrac{-1}{\rho(y^*)^{1/3}}\right) \\ \leq \;& \tfrac{2}{4^{4/3}\rho(y^*)^{5/6}}\left(\sqrt{\rho(y)} - \sqrt{\rho(y^*)}\right) \end{aligned}$$

where we used concavity of the function $u \mapsto -u^{-2/3}$. Then, we notice that

$$\frac{2}{4^{4/3}\rho(y^*)^{5/6}} = \frac{2}{3\sqrt{\rho(y^*)}} \cdot \frac{3}{4^{4/3}\rho(y^*)^{1/3}} = \frac{2}{3\sqrt{\rho(y^*)}}|f_{\text{hom}}(y^*)|$$

which yields the result as $f_{\text{hom}}(y^*) = f(x^*)$. □

Proposition 4.2 states that an approximative solution $x$ of (P) can be obtained as $s(y)y$, where $y$ is an approximate solution of the *homogenized* problem (H). We chose to minimize the square root of $\rho$ in (H) since, as we showed in Section 3, it has favorable properties for gradient methods. Let us now apply such scheme to (H).

## 4.2 Fast sublinear rates for squared uniform convexity

Motivated by Problem (H), let us study the gradient methods for solving problems of the form

$$\min_{y \in \mathcal{C}} g(y), \tag{4.3}$$

where, throughout this section, we assume the following:

- $\mathcal{C}$ is a closed convex set of some Euclidean space $E$,
- $g : E \mapsto \mathbb{R}$ is a convex function with $L$-Lipschitz continuous gradients and $g > 0$ on $\mathcal{C}$,
- $g^2$ satisfies the following uniform convexity property of degree 4:

$$g^2(x) - g^2(y) - \langle \nabla(g^2)(y), x - y\rangle \geq \frac{\mu^2}{4}\|x - y\|^4 \quad \forall x, y \in E. \tag{4.4}$$

Note that the assumptions on $g$ imply that $\mu \leq L$. Indeed, by Lipschitz gradient continuity,

$$g^2(x) \leq \left[g(y) + \langle \nabla g(y), x - y\rangle + \frac{L}{2}\|x - y\|^2\right]^2, \quad \forall x, y \in E.$$

Take now $x \to \infty$ with $y$ fixed. By comparing the dominant quartic term with that of the lower bound (4.4), we deduce that $\mu \leq L$.

In our problem of interest (H), we have $\mathcal{C} = \mathcal{K} \cap \{y : \langle y, c\rangle = 1\}$, the function $g = \sqrt{\rho}$ is convex and smooth with constant $L = 6\beta$ (Proposition 3.2), and $g^2 = \rho$ satisfies (4.4) with $\mu = \sqrt{4/3}\alpha$ (Proposition 3.3).

**Lemma 4.3.** *Denoting* $y^* = \operatorname{argmin}_{\mathcal{C}} g$ *we have for every* $y \in \mathcal{C}$

$$g^2(y) - g^2(y^*) \geq \frac{\mu^2}{4}\|y - y^*\|^4.$$

*Proof.* Let $y \in \mathcal{C}$. (4.4) implies

$$g^2(y) - g^2(y^*) - \langle \nabla(g^2)(y^*), y - y^* \rangle \geq \frac{\mu^2}{4}\|y - y^*\|^4.$$

Since $g \geq 0$, $y^*$ is also the minimizer of $g^2$ on $\mathcal{C}$, and therefore $\langle \nabla(g^2)(y^*), y - y^* \rangle \geq 0$. □

### 4.2.1 Gradient descent

We first study the convergence rate of the standard gradient descent method applied to (4.3): start with $y_0 \in E$ and iterate

$$y_{k+1} = \operatorname*{argmin}_{y \in \mathcal{C}} \left\{ \langle \nabla g(y_k), y - y_k \rangle + \frac{L}{2}\|y - y_k\|^2 \right\}, \quad k = 0, 1 \ldots, \tag{GD}$$

We define the *relative accuracy* (recall that $g(y^*) > 0$) at iteration $k$ as $\delta_k \triangleq \frac{g(y_k)}{g(y^*)} - 1$.

**Proposition 4.4.** *Let* $\delta > 0$. *Then the maximal number of iterations* $k$ *required for* (GD) *to reach a relative accuracy* $\delta_k \leq \delta$ *is at most*

$$\frac{12L}{\mu}\left( \ln_+(4\delta_0) + \left( \sqrt{\frac{1}{\delta}} - 2 \right)_+ \right).$$

*Proof.* Indeed, in view of our assumptions and Lemma 4.3, we have

$$\begin{aligned} g(y_{k+1}) &\leq \min_{y \in \mathcal{C}} \left\{ g(y_k) + \langle \nabla g(y_k), y - y_k \rangle + \tfrac{L}{2}\|y - y_k\|^2 \right\} \\ &\leq \min_{0 \leq \tau \leq 1} \left\{ g(y_k) + \tau \langle \nabla g(y_k), y^* - y_k \rangle + \tfrac{L\tau^2}{2}\|y^* - y_k\|^2 \right\} \\ &\leq \min_{0 \leq \tau \leq 1} \left\{ g(y_k) - \tau[g(y_k) - g(y^*)] + \tfrac{L\tau^2}{\mu}\sqrt{g^2(y_k) - g^2(y^*)} \right\}. \end{aligned}$$

The optimal value of $\tau$ in the latter problem is $\tau^* = \frac{\mu}{2L}\sqrt{\frac{g(y_k) - g(y^*)}{g(y_k) + g(y^*)}} \leq \frac{1}{2}$. Hence,

$$\delta_{k+1} \leq \delta_k - \frac{\mu}{4L}\delta_k \sqrt{\frac{g(y_k) - g(y^*)}{g(y_k) + g(y^*)}} = \delta_k - \frac{\mu}{4L}\frac{\delta_k^{3/2}}{\sqrt{2 + \delta_k}}.$$

Note that the values $\{\delta_k\}_{k \geq 0}$ decrease monotonically. Denote by $t_0$ the moment when $\delta_k \leq \frac{1}{4}$ for all $k > t_0$. Then, for all $k \leq t_0$ we have $2 \leq 8\delta_k$, and therefore

$$\delta_{k+1} \leq \left(1 - \frac{\mu}{12L}\right)\delta_k \leq \exp\left\{ -\frac{\mu}{12L} \right\}\delta_k, \quad 0 \leq k < t_0.$$

Hence, $t_0 \leq \frac{12L}{\mu}\ln_+(4\delta_0)$. For $k \geq t_0$, we have $\delta_{k+1} \leq \delta_k - \frac{\mu}{6L}\delta_k^{3/2}$. Therefore,

$$\delta_{k+1}^{-1/2} - \delta_k^{-1/2} = \frac{\delta_k - \delta_{k+1}}{\delta_k^{1/2}\delta_{k+1}^{1/2}(\delta_k^{1/2} + \delta_{k+1}^{1/2})} \geq \frac{\mu}{12L}.$$

Thus, we get $\delta_k^{-1/2} \geq \delta_{t_0}^{-1/2} + \frac{\mu}{12L}(k - t_0) \geq 2 + \frac{\mu}{12L}(k - t_0)$. □

### 4.2.2 Accelerated gradient descent with restarts

For a better convergence rate, consider the *accelerated gradient method* (see e.g., (Nesterov, 2020, Section 2.2.4))

$$\begin{aligned} &y_0 \in E, z_0 = y_0 \\ &y_{k+1} = \operatorname*{argmin}_{y \in \mathcal{C}} \left\{ \langle \nabla g(z_k), y - z_k \rangle + \frac{L}{2} \|y - z_k\|^2 \right\}, \\ &z_{k+1} = y_{k+1} + \frac{k}{k+3} \left( y_{k+1} - y_k \right), \end{aligned} \tag{AGD}$$

for $k \geq 0$. Let us write $\mathrm{AGD}(y_0, k)$ the output $y_k$ of this algorithm initialized at $y_0$ with $k$ iterations. Using the technique of scheduled restarts (Nemirovski and Yudin, 1983; Nesterov, 2020) allows to prove an improved convergence rate for our setting.

**Proposition 4.5.** *For $t \geq 0$ and $y_0 \in E$, consider the following scheme:*

$$\begin{aligned} &\textit{Compute } \hat{y}_{t+1} = \mathrm{AGD}(y_t, 2^t), \\ &\textit{Choose } y_{t+1} = \operatorname*{argmin}_{y} \left\{ g(y) \,:\, y \in \{y_0, \dots, y_t, \hat{y}_{t+1}\} \right\}, \end{aligned} \tag{AGD-restart}$$

*Then the number of total iterations (i.e., gradient oracle calls in (AGD)) needed to reach a relative accuracy less than $\delta$ is at most*

$$16 \sqrt{\frac{L}{\mu}} \frac{\left( \delta_0^{1/4} + 2^{1/4} \right)}{\delta^{1/4}}$$

*Proof.* Denote $\delta_t = \frac{g(y_t)}{g(y^*)} - 1$. Let us prove by induction the following bound:

$$\delta_t \leq \frac{\delta_0}{16^{t - t0}} \quad \forall t \geq t_0, \tag{4.8}$$

where $t_0$ is the smallest integer such that

$$4^{t_0} \geq \frac{64 L}{\mu} \left( 1 + \sqrt{\frac{2}{\delta_0}} \right). \tag{4.9}$$

Condition (4.8) is satisfied for $t = t_0$, because the outer scheme (AGD-restart) is monotone by construction. Assume that it holds for some $t \geq t_0$. Then, by the known convergence result of the (AGD) scheme (Beck, 2017, Thm 10.34), we have

$$g(y_{t+1}) - g(y^*) \leq \frac{2L \|y_t - y^*\|^2}{(2^t)^2}.$$

Then,by Lemma 4.3, we get the following bound:

$$\delta_{t+1} \leq \frac{4L}{4^t \mu} \sqrt{(1 + \delta_t)^2 - 1} \leq \frac{4L}{4^t \mu} \left( \delta_t + \sqrt{2 \delta_t} \right) \leq \frac{4L}{4^t \mu} \left( \frac{\delta_0}{16^{t - t_0}} + \frac{\sqrt{2 \delta_0}}{4^{t - t_0}} \right)$$

where the last line uses the induction hypothesis (4.8). It follows that

$$\begin{aligned}\delta_{t+1} &\leq \frac{\delta_0}{16^{t+1-t_0}} \cdot \frac{4 \cdot 16L}{4^t \mu} \left(1 + \sqrt{\frac{2}{\delta_0}} 4^{t-t_0}\right) \\ &\leq \frac{\delta_0}{16^{t+1-t_0}} \cdot \frac{64L}{4^{t_0} \mu} \left(4^{t_0-t} + \sqrt{\frac{2}{\delta_0}}\right) \\ &\leq \frac{\delta_0}{16^{t+1-t_0}} \cdot \frac{64L}{4^{t_0} \mu} \left(1 + \sqrt{\frac{2}{\delta_0}}\right) \leq \frac{\delta_0}{16^{t+1-t_0}},\end{aligned}$$

where the last inequality follows from (4.9), proving thus the induction.

Now, assume that after $T$ outer iterations, procedure (AGD-restart) has reached an accuracy $\delta_T \geq \delta$ for some $\delta > 0$. Hence, by (4.8) we have $16^T \leq 16^{t_0} \delta_0 / \delta$. Note also by definition (4.9) of $t_0$,

$$4^{t_0 - 1} \leq \frac{64L}{\mu} \left(1 + \sqrt{\frac{2}{\delta_0}}\right).$$

Thus, we can bound the total number of inner iterations performed by the method as

$$\begin{aligned}\sum_{t=0}^{T-1} 2^t = 2^T - 1 &\leq 2^{t_0} \left(\frac{\delta_0}{\delta}\right)^{1/4} \leq 2^{t_0} \left(\frac{\delta_0}{\delta}\right)^{1/4} \\ &\leq \sqrt{\frac{2^8 L}{\mu} \left(1 + \sqrt{\frac{2}{\delta_0}}\right)} \left(\frac{\delta_0}{\delta}\right)^{1/4} \\ &\leq 16 \sqrt{\frac{L}{\mu}} \left(1 + \left(\frac{2}{\delta_0}\right)^{1/4}\right) \left(\frac{\delta_0}{\delta}\right)^{1/4}.\end{aligned}$$

□

### 4.3 Convergence rate on the quartic problem

After studying the complexity of methods for solving (H), we get back to the original problem (P): the procedure is summarized in Algorithms 1 (basic version) and 2 (accelerated version).

**Implementation of the projected gradient step** We describe how to compute the projected gradient step to get $y_{k+1}$ in Algorithm 1. First, note that we can decompose it as a gradient step followed with a projection step:

$$\begin{aligned}\hat{y}_{k+1} &= y_k - \frac{1}{12\beta\sqrt{\rho(y_k)}} B^{-1} \nabla \rho(y_k), \\ y_{k+1} &= \operatorname*{argmin}_{\substack{y \in \mathcal{K} \\ \langle y, c \rangle = 1}} \frac{1}{2} \|y - \hat{y}_{k+1}\|^2,\end{aligned} \tag{4.10}$$

where we recall that $B$ is the positive definite operator defining the norm.

When the original problem is unconstrained ($\mathcal{K} = E$), the projection is

$$y_{k+1} = \hat{y}_{k+1} + \left(\frac{1 - \langle \hat{y}_{k+1}, c \rangle}{\|c\|_*^2}\right) B^{-1} c.$$

When $\mathcal{K}$ is a cone to which the projection is easily computable, we can solve a dual of Problem

(4.10) which writes

$$\max_{t\in\mathbb{R}} \min_{y\in\mathcal{K}} \left\| y - \hat{y}_{k+1} - \frac{t}{2} B^{-1} c \right\|^2 + t\left(1 - \langle \hat{y}_{k+1}, c\rangle\right) - \frac{1}{2} t^2 \|c\|_*^2.$$

The solution $t^*$ can be computed with a bisection method on $t$, which requires only a logarithmic number of projections on $\mathcal{K}$. Then, the projection is $y_{k+1} = \hat{y}_{k+1} + t^* B^{-1} c$.

**Algorithm 1** Homogenized-GD$(x_0, \beta, K)$

**Initialize** $y_0 \in E$.
**for** $k = 0, 1, \ldots, K-1$ **do**
  $y_{k+1} = \operatorname{argmin}\left\{ \frac{1}{2\sqrt{\rho(y_k)}} \langle \nabla\rho(y_k), y - y_k\rangle + 3\beta\|y - y_k\|^2 \,:\, y \in \mathcal{K},\, \langle c, y\rangle = 1 \right\}$.
**end for**
**Return** $x_K = y_K / \left(4\rho(y_K)\right)^{1/3}$.

**Algorithm 2** Accelerated Homogenized-GD$(x_0, \beta, T)$

**Initialize** $y_0 \in E$.
**for** $t = 0, 1, \ldots, T-1$ **do**
  Set $y_t^0 = z_t^0 = y_{t-1}$
  **for** $k = 0, 1, \ldots, 2^t - 1$ **do**
    $y_t^{k+1} = \operatorname{argmin}\left\{ \frac{1}{2\sqrt{\rho(z_t^k)}} \langle \nabla\rho(z_t^k), y - z_t^k\rangle + 3\beta\|y - z_t^k\|^2 \,:\, y \in \mathcal{K},\, \langle c, y\rangle = 1 \right\}$.
    $z_t^{k+1} = y_t^{k+1} + \frac{k}{k+3}\left(y_t^{k+1} - y_t^k\right)$
  **end for**
  Set $y_t = \operatorname{argmin}_y \left\{\rho(y) : y \in \{y_{t-1}, y_t^{2^t}\}\right\}$
**end for**
**Return** $x_T = y_T / \left(4\rho(y_T)\right)^{1/3}$.

**Theorem 4.6.** *Algorithm 1 finds a point $x_K$ satisfying*

$$\frac{f(x_K) - f(x^*)}{|f(x^*)|} \le \delta$$

*in at most $\mathcal{O}\left(\kappa\left(\log(\frac{\delta_0}{\delta}) + \sqrt{\frac{1}{\delta}}\right)\right)$ iterations, where the initial relative accuracy is $\delta_0 = \sqrt{\frac{\rho(y_0)}{\rho(y^*)}} - 1$, with $y^*$ being the solution of the homogenized problem (H), and $\kappa = \beta/\alpha$ is the quartic condition number defined by (1.1).*

*Proof.* First, note that the algorithm outputs a point $y = s(y)y$ where $s$ is defined in (4.1). Then, Proposition 4.2 ensures

$$\frac{f(x) - f(x^*)}{|f(x^*)|} \le \frac{2}{3} \frac{\sqrt{\rho(y)} - \sqrt{\rho(y^*)}}{\sqrt{\rho(y^*)}}, \tag{4.11}$$

where $y^*$ is a solution of the homogenized problem (H). This problem satisfies all assumptions for the general class studied in Section 4.2: $g = \sqrt{\rho}$ is convex and smooth with constant $L = 6\beta$ (Proposition 3.2), and $g^2 = \rho$ satisfies (4.4) with $\mu = \sqrt{\frac{4}{3}}\alpha$ (Proposition 3.3). Algorithm 1 corresponds to procedure (GD) applied to (H).

Proposition 4.4 states that the number of iterations $K$ needed for finding a point $y_K$ satisfying $\sqrt{\rho(y_K)} \leq (1+\delta)\sqrt{\rho(y^*)}$ is at most

$$\frac{12L}{\mu}\left(\ln_+(4\delta_0) + \left(\sqrt{\frac{1}{\delta}} - 2\right)_+\right).$$

Recalling that $L = 6\beta, \mu = \sqrt{\frac{4}{3}}\alpha$, we have $\frac{L}{\mu} = \sqrt{27}\kappa$, which, combined with (4.11), allows to conclude. □

A similar reasoning with Proposition 4.5 justifies the accelerated variant.

**Theorem 4.7.** *Accelerated homogenized gradient descent (Algorithm 2) finds a point $x_K$ satisfying*

$$\frac{f(x_K) - f(x^*)}{|f(x^*)|} \leq \delta$$

*in at most $\mathcal{O}\left(\frac{\sqrt{\kappa}}{\delta^{1/4}}\left(1 + \delta_0^{1/4}\right)\right)$ projected gradient steps, where $\delta_0 = \sqrt{\frac{\rho(y_0)}{\rho(y^*)}} - 1$.*

## 5 Optimal preconditioning

We showed that the efficiency of our methods crucially depends on the quartic condition number. In this section, we propose to search for a norm-inducing operator $B$, or *preconditioner*, such that the condition number of $\rho$ with respect to the Euclidean norm $\|\cdot\|_B$ is small.

We focus on the particular case when $\rho$ corresponds to the quartic part of a quadratic inverse problem (QIP) with positive semidefinite operators:

$$\rho(x) = \sum_{i=1}^{m} \langle x, B_i x\rangle^2. \tag{5.1}$$

For simplicity, we choose $E = \mathbb{R}^n$, and let the following standing assumptions hold.

**Assumption 5.1.** *The $m$-uple of operators $\mathcal{B} = (B_1, \dots B_m) \in (\mathbb{S}^n)^m$ is such that*

*(i)* $m \geq n$,

*(ii)* $B_1, \dots B_m$ *are nonzero positive semidefinite matrices,*

*(iii) each* $B_i$ *has rank at most* $r$, *where* $1 \leq r \leq n$,

*(iv) the matrix* $\sum_{i=1}^m B_i$ *is of full rank.*

These assumptions are satisfied for the phase retrieval problem (2.2), where $r = 2$, and distance matrix completion, for which $r = 3$ in the majority of applications[2].

We start in Section 5.1 by addressing a more general goal. That is a quadratic approximation of the function $x \mapsto (\sum_{i=1}^m \langle B_i x, x\rangle^p)^{1/p}$ for some general power $p > 1$. We show that there exists

[2]The full rank assumption holds generically for the phase retrieval with $m$ high enough, and is satisfied for distance matrix completion if the graph is connected.

an operator $B_*$ for which the approximation ratio of this function with respect to $\|\cdot\|^2_{B_*}$ is $n^{\frac{p-1}{p}}$, and that it can be efficiently computed by a fixed-point iteration method described in Section 5.2.

### 5.1 Existence and characterization of the optimal preconditioner

Let us fix a power $p > 1$, and write its conjugate power $q$ as $\frac{1}{p} + \frac{1}{q} = 1$. For a given $m$-uple of matrices $\mathcal{B} = (B_1, \dots B_m)$ satisfying the assumptions above, we define the function

$$W_{\mathcal{B},p}(x) = \sum_{i=1}^{m} \langle B_i x, x\rangle^p,$$

for which the quartic case corresponds to $p = 2$. Consider the following problem.

*How good can be a quadratic approximation of the function $W^{1/p}_{\mathcal{B},p}(\cdot)$?*

Let us look at quadratic function $\langle B(\tau)\cdot,\cdot\rangle$ with an operator $B$ of the following form:

$$B(\tau) = \sum_{i=1}^{m} \tau_i B_i, \quad \tau \in \mathbb{R}^m_{++}.$$

Since $\tau > 0$, Assumption 5.1 implies that $B(\tau) \succ 0$. It induces therefore a norm $\|x\|_{B(\tau)} = \langle B(\tau)x, x\rangle^{\frac{1}{2}}$. Let us look how good is it for approximating $W_{\mathcal{B},p}$.

#### 5.1.1 Comparing $W_{\mathcal{B},p}(x)$ and $\|x\|_{B(\tau)}$

Our analysis relies on the quantities

$$\ell_i(\tau; \mathcal{B}) = \mathrm{Tr}\left[\tau_i B(\tau)^{-1} B_i\right], \tag{5.2}$$

defined for $i = 1 \dots m$ and $\tau \in \mathbb{R}^m_{++}$. The $\ell_i(\tau;\mathcal{B})$ measures the importance of the matrix $\tau_i B_i$ in forming the whole $B(\tau)$. If the $B_i$'s are of rank 1, they correspond to the classical notion of *leverage scores* used in regression analysis and randomized Linear Algebra (Mahoney, 2011; Martinsson and Tropp, 2020).

**Lemma 5.2.** *Under Assumption 5.1, we have for any $\tau \in \mathbb{R}^m_{++}$ and $i = 1 \dots m$*

*(i)* $\ell_i(\tau;\mathcal{B}) \in [0, \mathrm{rank}(B_i)]$,

*(ii)* $\sum_{j=1}^m \ell_j(\tau;\mathcal{B}) = n$,

*(iii)* $\tau_i\langle x, B_i x\rangle \le \ell_i(\tau;\mathcal{B})\,\|x\|^2_{B(\tau)}$ *for all* $x \in \mathbb{R}^n$.

*Proof.* Since all $B_i \succeq 0$, we have $\tau_i B_i \preceq B(\tau)$ and therefore the matrix

$$Z_i = \tau_i B(\tau)^{-\frac{1}{2}} B_i B(\tau)^{-\frac{1}{2}}$$

satisfies $0 \preceq Z_i \preceq I_n$. This implies $0 \le \mathrm{Tr}\,[Z_i] \le \mathrm{rank}(Z_i) = \mathrm{rank}(B_i)$, proving the Item (i) since $\ell_i(\tau;\mathcal{B}) = \mathrm{Tr}\,[Z_i]$. Item (ii) follows from the equality

$$\sum_{i=1}^{m} \mathrm{Tr}\left[\tau_i B(\tau)^{-1} B_i\right] = \mathrm{Tr}\left[B(\tau)^{-1} B(\tau)\right] = n.$$

Finally, to prove (iii) we write

$$\begin{aligned}\langle B_i x, x\rangle &= \|B_i^{\frac12} x\|_2^2 = \|B_i^{\frac12} B(\tau)^{-\frac12} B(\tau)^{\frac12} x\|_2^2 \\ &\le \|B_i^{\frac12} B(\tau)^{-\frac12}\|_F^2 \, \|B(\tau)^{\frac12} x\|_2^2 = \operatorname{Tr}\left[B(\tau)^{-\frac12} B_i B(\tau)^{-\frac12}\right] \langle Bx, x\rangle \\ &= \tau_i^{-1} \ell_i(\tau;\mathcal{B}) \, \|x\|_{B(\tau)}^2.\end{aligned}$$

□

Using the quantities $\ell_i(\tau;\mathcal{B})$, we can compare $\|x\|_{B(\tau)}^2$ with $W_{\mathcal{B},p}(x)$.

**Proposition 5.3.** *For any $\tau \in \mathbb{R}^m_{++}$ and $x \in \mathbb{R}^n$ we have*

$$\left(\frac{1}{\|\tau\|_q}\right) \|x\|_{B(\tau)}^2 \le W_{\mathcal{B},p}(x)^{\frac1p} \le \left(\max_{i=1\dots m} \frac{\ell_i(\tau;\mathcal{B})^{\frac1q}}{\tau_i}\right) \|x\|_{B(\tau)}^2. \tag{5.3}$$

*Proof.* The lower bound follows from Hölder's inequality, as $\frac1p + \frac1q = 1$:

$$\|x\|_{B(\tau)}^2 = \sum_{i=1}^m \tau_i \langle x, B_i x\rangle \le \left(\sum_{i=1}^m \tau_i^q\right)^{\frac1q} \left(\sum_{i=1}^m \langle x, B_i x\rangle^p\right)^{\frac1p} = \|\tau\|_q \, W_{\mathcal{B},p}(x)^{\frac1p}.$$

For the upper bound, we write

$$\begin{aligned}W_{\mathcal{B},p}(x) &= \sum_{i=1}^m \langle x, B_i x\rangle^p = \sum_{i=1}^m \frac{(\tau_i \langle x, B_i x\rangle)^{p-1}}{\tau_i^p} \tau_i \langle x, B_i x\rangle \\ &\le \left(\max_{i=1\dots m} \frac{(\tau_i \langle x, B_i x\rangle)^{p-1}}{\tau_i^p}\right) \|x\|_{B(\tau)}^2 \le \left(\max_{i=1\dots m} \frac{\ell_i(\tau;\mathcal{B})^{p-1}}{\tau_i^p}\right) \|x\|_{B(\tau)}^{2p},\end{aligned}$$

where the last inequality follows from Lemma 5.2(iii). □

### 5.1.2 Existence of $n^{\frac{p-1}{p}}$-quadratic approximation

Proposition 5.3 gives an upper bound for the condition number of $W_{\mathcal{B},p}^{1/p}$ with respect to $\|\cdot\|_{B(\tau)}$. We now seek for the d coefficients $\tau$ that minimize this upper bound. Since the condition number is invariant by rescaling of $\tau$, we can choose to impose the condition $\|\tau\|_q = 1$.

Let us show that the factor in the right-hand side of (5.3) is at least $n^{\frac1q}$.

**Proposition 5.4.** *For every $\tau \in \mathbb{R}^m_{++}$ such that $\|\tau\|_q = 1$, we have*

$$\max_{i=1\dots m} \frac{\ell_i(\tau;\mathcal{B})^{\frac1q}}{\tau_i} \ge n^{\frac1q}.$$

*Proof.* If, on the contrary, we have $\ell_i(\tau;\mathcal{B}) < n\tau_i^q$ for every $i \in \{1\dots m\}$, we reach a contradiction as $\sum_{i=1}^m \ell_i(\tau;\mathcal{B}) = n$ by Lemma 5.2(ii) and $\sum_{i=1}^m \tau_i^q = 1$. □

It appears that there is a unique vector $\tau^*$ which attains this minimal value.

**Proposition 5.5.** *There exists a unique vector $\tau^* \in \mathbb{R}^m_{++}$ such that*

$$\frac{\ell_i(\tau^*;\mathcal{B})^{\frac1q}}{\tau_i^*} = n^{\frac1q}, \quad i = 1\dots m. \tag{5.4}$$

*This vector satisfies $\|\tau^*\|_q = 1$. It is a unique solution to the minimization problem*

$$\min_{\tau \in \mathbb{R}^m} V(\tau) = -\ln\det B(\tau) + \frac{n}{q}\|\tau\|_q^q. \tag{5.5}$$

*Proof.* As $q > 1$, the function $V$ is strictly convex and differentiable on the set

$$\mathcal{D} = \{\tau \in \mathbb{R}^m \,:\, B(\tau) \succ 0\}.$$

Since $V(\tau) \to \infty$ as $\tau \to \partial\mathcal{D}$ (the boundary of $\mathcal{D}$) or $\|\tau\|_q \to \infty$, Problem (5.5) admits a unique minimizer $\tau^*$, which belongs to $\mathcal{D}$.

As the gradient of function $B \mapsto \ln\det B$ is $B^{-1}$, the first-order optimality conditions write

$$\operatorname{Tr}\left[B(\tau^*)^{-1}B_i\right] = n\operatorname{sign}(\tau_i^*)|\tau_i^*|^{q-1}, \quad i = 1\dots m. \tag{5.6}$$

As we assumed the $B_i$'s to be nonzero and $\tau^* \in \mathcal{D}$, we have $\operatorname{Tr}\left[B(\tau^*)^{-1}B_i\right] > 0$ and therefore $\tau_i^* > 0$ for every $1 \le i \le m$. Now, as the left-hand side of (5.6) is $\ell_i(\tau^*;\mathcal{B})/\tau_i^*$, it follows that

$$\ell_i(\tau^*;\mathcal{B}) = n(\tau_i^*)^q, \quad i = 1\dots m,$$

which proves (5.4). Summing up these equations and using Lemma 5.2(ii), we get

$$\|\tau^*\|_q^q = \frac{1}{n}\sum_{i=1}^n \ell_i(\tau^*;\mathcal{B}) = 1.$$

□

In the case $r = 1$, vector $\tau^*$ corresponds exactly to the *Lewis weights* (Lewis, 1978; Cohen and Peng, 2015) used in $\ell_p$ subspace sampling. We allow operators $B_1, \dots B_m$ to have a higher rank, and show that the same construction can be used for building an efficient quadratic approximation for $W_{\mathcal{B},p}^{1/p}(\cdot)$. We summarize the discussion in the following theorem.

**Theorem 5.6** (Optimal quadratic approximation of $W_{\mathcal{B},p}$)**.** *For every $\mathcal{B} = (B_1, \dots B_m)$ satisfying Assumption 5.1, there exists a matrix $B_* \in \mathbb{S}^n_{++}$ such that for every $x \in \mathbb{R}^n$*

$$\|x\|_{B_*}^2 \le \left(\sum_{i=1}^m \langle B_i x, x\rangle^p\right)^{\frac{1}{p}} \le n^{1-\frac{1}{p}}\|x\|_{B_*}^2 \tag{5.7}$$

*Proof.* This follows by taking $B_* = B(\tau^*)$, where $\tau^*$ is given by Proposition 5.5, and applying Proposition 5.3. □

### 5.1.3 Optimality for general $\mathcal{B}$

One can ask if the constant $n^{1-1/p}$ in (5.7) is optimal. This is true for a *generic* $\mathcal{B}$, although it might be better for specific choices.

Indeed, let us choose $m = n$ and $B_i = e_i e_i^T$ for $i = 1\dots n$, where $\{e_i\}_{1\le n}$ denote the coordinate vectors of $\mathbb{R}^n$. We have

$$W_{\mathcal{B},p}(x) = \sum_{i=1}^n x_i^{2p} = \|x\|_{2p}^{2p}.$$

A symmetry argument implies that the best quadratic approximations $\langle B\cdot,\cdot\rangle$ are isotropic, i.e. $B \propto I_n$. For such $B$, the approximation ratio is $n^{1-1/p}$ as we have

$$n^{\frac{1}{p}-1}\|x\|_2^2 \le \|x\|_{2p}^2 \le \|x\|_2^2,$$

which is tight (take $x = (1 \dots 1)$ for matching the lower bound and $x = e_1$ for the upper bound). This shows that the bound (5.7) is not improvable for general $\mathcal{B}$.

## 5.2 Fixed-point algorithm for computing the optimal weights

We now show how the coefficients $\tau^*$ can be computed using a fixed point algorithm when $p < 2$. This process converges only for $p < 2$, but we will show in Section 5.2.2 that we can also deal with the case $p = 2$ (which corresponds to the quartic applications) by using the fixed-point method with a surrogate power $p' < 2$.

### 5.2.1 Case $p < 2$

The equations (5.4), which characterize $\tau^*$, are as follows:

$$\ell_i(\tau_i; \mathcal{B}) = n\tau_i^{\frac{p}{p-1}}, \quad i = 1 \dots m. \tag{5.8}$$

By definition (5.2) of $\ell_i$, (5.8) can be written equivalently as the fixed-point equation

$$\tau_i = \left(\frac{1}{n}\mathrm{Tr}\left[B(\tau)^{-1}B_i\right]\right)^{p-1}, \quad i = 1 \dots m.$$

Based on this observation, we propose in Algorithm 3 a fixed-point method for approximating $\tau^*$. This method was proposed initially by (Cohen and Peng, 2015) for computing the Lewis weights with $r = 1$, and we generalize it to possibly larger rank $r$.

**Algorithm 3** Computing approximate optimal weights of order $p \in (1, 2)$

1: **Input** $p \in (1, 2)$, $\mathcal{B} = (B_1, \dots B_m)$, precision $\epsilon > 0$.

2: **Initialize** $\tau^{(0)} = (m^{\frac{1}{p}-1}, \dots m^{\frac{1}{p}-1})$.

3: **for** $k = 0, 1, \dots$ do

4: $\quad \tau_i^{(k+1)} = \left(\frac{1}{n}\mathrm{Tr}\left[\left(\sum_{j=1}^m \tau_j^{(k)} B_j\right)^{-1} B_i\right]\right)^{p-1}$ for $i = 1 \dots m$,

5: **until** $D_L(\tau^{(k+1)}, \tau^{(k)}) \le \frac{2-p}{p-1}\epsilon$.

**Convergence analysis of Algorithm 3.** Define the fixed-point operator $T_p$ as

$$[T_p(\tau)]_i = \left(\frac{1}{n}\mathrm{Tr}\left[B(\tau)^{-1}B_i\right]\right)^{p-1}, \quad \tau \in \mathbb{R}^m_{++}, \quad i = 1 \dots m. \tag{5.9}$$

Note that if $\tau > 0$, then $B(\tau)$ is positive definite and $T_p(\tau) > 0$. Therefore, $T_p$ is a map from $\mathbb{R}^m_{++}$ to itself. We define the log-infinity distance $D_L(u, v)$ for $u, v \in \mathbb{R}^m_{++}$:

$$D_L(u, v) = \max_{i=1\dots m} |\ln u_i - \ln v_i|.$$

We show that, for $p < 2$, $T_p$ is a contraction with respect to the distance $D_L$.

**Lemma 5.7.** *For every $\tau, \tilde{\tau} \in \mathbb{R}^m_{++}$ we have $D_L\Big(T_p(\tau), T_p(\tilde{\tau})\Big) \le (p-1)\, D_L(\tau, \tilde{\tau})$.*

*Proof.* Let $\tau, \tilde{\tau} \in \mathbb{R}^m_{++}$, and let $\Delta = \exp\left[D_L(\tau, \tilde{\tau})\right]$ so that

$$\frac{1}{\Delta}\tilde{\tau}_i \leq \tau_i \leq \Delta\tilde{\tau}_i, \quad i = 1 \dots m.$$

This implies that $\Delta^{-1}\tilde{\tau}_i B_i \preceq \tau_i B_i \preceq \Delta\tilde{\tau}_i B_i$ for every $i = 1 \dots m$ and therefore

$$\Delta^{-1} B(\tilde{\tau})^{-1} \preceq B(\tau)^{-1} \preceq \Delta B(\tilde{\tau})^{-1}.$$

Using the expression (5.9) for $T_p$ yields, recalling that $p > 1$,

$$\Delta^{-(p-1)}\left[T_p(\tilde{\tau})\right]_i \leq \left[T_p(\tau)\right]_i \leq \Delta^{p-1}\left[T_p(\tilde{\tau})\right]_i, \quad i = 1 \dots m,$$

which implies $D_L\left(T_p(\tau), T_p(\tilde{\tau})\right) \leq (p-1)\ln\Delta$. □

The contraction property allows to establish the convergence rate to the fixed point.

**Proposition 5.8.** *Assume that $1 < p < 2$. The iterates of the fixed-point process*

$$\begin{cases} \tau^{(0)} = \left(m^{\frac{1}{p}-1}, \dots, m^{\frac{1}{p}-1}\right), \\ \tau^{(k+1)} = T_p(\tau^{(k)}), \quad k = 0, 1 \dots \end{cases}$$

*converge to the solution $\tau^*$ of (5.4) and satisfy for $k \geq 1$*

$$D_L\left(\tau^{(k)}, \tau^*\right) \leq \left(1 - \frac{1}{p}\right)^{k+1} \ln(rm).$$

*Proof.* As, by Lemma 5.7, $T_p$ is a contraction for $p < 2$, and $\tau^*$ is the unique vector satisfying $T_p(\tau^*) = \tau^*$, we can apply the Banach fixed point theorem in the space $(\mathbb{R}^m_{++}, D_L)$, which proves convergence. Moreover, for $k \geq 1$, we have

$$D_L\left(\tau^{(k)}, \tau^*\right) \leq (p-1)^{k-1} D_L\left(\tau^{(1)}, \tau^*\right), \tag{5.10}$$

To bound the distance of $\tau^{(1)}$ to the optimum, let us show that $B(\tau^{(0)})$ and $B(\tau^*)$ are comparable. Since $\|\tau^{(0)}\|_q = 1$, Proposition 5.3 as applied to $\tau^{(0)}$ yields

$$\|x\|^2_{B(\tau^{(0)})} \leq W_{\mathcal{B},p}(x)^{\frac{1}{p}} \leq \left(\max_{i=1\dots m} m\,\ell_i(\tau^{(0)}; \mathcal{B})\right)^{1-\frac{1}{p}} \|x\|^2_{B(\tau^{(0)})} \tag{5.11}$$

for every $x \in \mathbb{R}^n$. By Lemma 5.2(i), we have

$$\max_{i=1\dots m} \ell_i(\tau^{(0)}; \mathcal{B}) \leq \max_{i=1\dots m} \operatorname{rank}(B_i) = r. \tag{5.12}$$

Similarly, according to Theorem 5.6, $B(\tau^*)$ satisfies

$$\|x\|^2_{B(\tau^*)} \leq W_{\mathcal{B},p}(x)^{\frac{1}{p}} \leq n^{1-\frac{1}{p}} \|x\|^2_{B(\tau^*)}. \tag{5.13}$$

Combining (5.11), (5.12) and (5.13), we get

$$(rm)^{-(1-\frac{1}{p})} B(\tau^*) \preceq B(\tau^{(0)}) \preceq n^{1-\frac{1}{p}} B(\tau^*).$$

This implies that for every $i = 1 \dots m$,

$$n^{-\frac{(p-1)^2}{p}}\left[T_p(\tau^*)\right]_i \leq \left[T_p(\tau^{(0)})\right]_i \leq (rm)^{\frac{(p-1)^2}{p}}\left[T_p(\tau^*)\right]_i$$

and therefore

$$D_L\big(T_p(\tau^{(1)}), \tau^*\big) \leq p^{-1}(p-1)^2 \ln\left[\max\left(n, rm\right)\right] \leq p^{-1}(p-1)^2 \ln(rm),$$

as we assumed $m \geq n$. Combining with (5.10) achieves the result.

□

This result allows us to estimate the number of iterations needed to compute an approximate solution to (5.4), and therefore the complexity of Algorithm 3.

**Proposition 5.9.** *Assume that $1 < p < 2$ and let $\epsilon > 0$. Then, Algorithm 3 with precision $\epsilon$ performs at most*

$$\frac{p-1}{2-p}\left[\ln \frac{\ln(rm)}{(2-p)\epsilon}\right]_+$$

*iterations, and its output $\overline{\tau}$ satisfies*

$$e^{-\epsilon}\,\|x\|^2_{B(\overline{\tau})} \leq \left(\sum_{i=1}^m \langle B_i x, x\rangle^p\right)^{\frac{1}{p}} \leq e^{\epsilon}\, n^{1-\frac{1}{p}}\,\|x\|^2_{B(\overline{\tau})}. \tag{5.14}$$

*Proof.* By Proposition 5.8, we can bound the distance between successive iterates:

$$D_L\left(\tau^{(k)}, \tau^{(k-1)}\right) \leq D_L\left(\tau^{(k)}, \tau^*\right) + D_L\left(\tau^{(k-1)}, \tau^*\right) \leq (p-1)^k \ln(rm).$$

Since Algorithm 3 stops as soon as $D_L\left(\tau^{(k)}, \tau^{(k-1)}\right) \leq \frac{2-p}{p-1}\epsilon$, the number $k$ of the iterations performed satisfies the bound $(p-1)^{k+1}\ln(rm) \leq (2-p)\epsilon$. Hence

$$\frac{\ln(rm)}{(2-p)\epsilon} \leq \left(\frac{1}{p-1}\right)^{k+1} \leq \exp\left\{\frac{(k+1)(2-p)}{p-1}\right\}.$$

To analyse the approximation quality of the last iterate $\overline{\tau} = \tau^{(K)}$, we use the contraction property of $T_p$ and the stopping criterion:

$$\begin{aligned} D_L\left(\tau^{(K)}, \tau^*\right) &\leq D_L\left(\tau^{(K+1)}, \tau^{(K)}\right) + D_L\left(\tau^{(K+1)}, \tau^*\right) \\ &\leq (p-1) D_L\left(\tau^{(K)}, \tau^{(K-1)}\right) + (p-1) D_L\left(\tau^{(K)}, \tau^*\right) \\ &\leq (2-p)\epsilon + (p-1) D_L\left(\tau^{(K)}, \tau^*\right). \end{aligned}$$

Hence, $e^{-\epsilon}B(\tau^*) \preceq B(\tau^{(K)}) \preceq e^{\epsilon}B(\tau^*)$, and using the property of $B(\tau^*)$ from Theorem 5.6, we get (5.14). □

**Computational complexity.** In practice, we usually have an access to a compact representation of $B_1 \dots B_m$ as $B_i = U_i U_i^T$, where $U_i \in \mathbb{R}^{n\times r}$ (in our examples, such representation is explicit). Then, the quantity $\ell_i(\tau)$ can be computed as

$$\ell_i(\tau) = \tau_i \mathrm{Tr}\left[U_i^T B(\tau)^{-1} U_i\right],$$

which takes $\mathcal{O}(n^2 r)$ operations for each $i = 1 \dots m$. Forming and inverting the matrix $B(\tau)$ requires $\mathcal{O}(mn^2)$ operations. Therefore, the total complexity of one iteration of Algorithm 3 is $\mathcal{O}(mrn^2)$, and we need to perform $\mathcal{O}\left(\ln(rm)/(2-p)\right)$ such iterations as we are satisfied with a constant factor approximation of the optimal preconditioner.

### 5.2.2 Quartic case ($p = 2$)

The fixed-point method from the previous section converges only for $p < 2$. However, we are particularly interested in computing the weights $\tau^*$ for $p = 2$, which corresponds to preconditioning

of quartic functions. In this case, let us show that a constant factor approximation of the optimal $\sqrt{n}$ ratio can be obtained by the previous method with a well-chosen surrogate power $p' < 2$.

**Corollary 5.10.** *Let $\omega > 1$. The Algorithm 3 with precision $\epsilon > 0$ and power*

$$p' = \frac{2 \ln \frac{m}{n}}{\ln \frac{m}{n} + 2 \ln \omega}$$

*performs at most $\mathcal{O}\left([\ln \ln(rm) - \ln \ln \omega - \ln \epsilon]_+ \left(\ln \frac{m}{n} / \ln \omega\right)\right)$ iterations, and the output $\overline{\tau}$ satisfies*

$$e^{-\epsilon} \left( \frac{1}{m^{\frac{1}{p'} - \frac{1}{2}}} \|x\|^2_{B(\overline{\tau})} \right) \leq \left( \sum_{i=1}^m \langle B_i x, x \rangle^2 \right)^{\frac{1}{2}} \leq e^{\epsilon} \omega \sqrt{n} \left( \frac{1}{m^{\frac{1}{p'} - \frac{1}{2}}} \|x\|^2_{B(\overline{\tau})} \right).$$

*Proof.* As $p' < 2$, the classical inequality between $\ell_2$ and $\ell_{p'}$ norms yields

$$\frac{1}{m^{\frac{1}{p'} - \frac{1}{2}}} W_{\mathcal{B},p'}(x)^{\frac{1}{p'}} \leq W_{\mathcal{B},2}(x)^{\frac{1}{2}} \leq W_{\mathcal{B},p'}(x)^{\frac{1}{p'}}.$$

Combining with inequality (5.14) from Proposition 5.9 gives

$$\frac{1}{m^{\frac{1}{p'} - \frac{1}{2}}} e^{-\epsilon} \|x\|^2_{B(\overline{\tau})} \leq W_{\mathcal{B},2}(x)^{\frac{1}{2}} \leq e^{\epsilon} n^{1 - \frac{1}{p'}} \|x\|^2_{B(\overline{\tau})},$$

and we conclude by noticing that $p'$ was chosen so that $\left(\frac{m}{n}\right)^{\frac{1}{p'} - \frac{1}{2}} = \omega$. To estimate the total number of iterations, we use Proposition 5.9 again and write

$$\frac{1}{2 - p'} = \frac{1}{4} + \frac{\ln \frac{m}{n}}{4 \ln \omega} = \mathcal{O}\left( \frac{\ln \frac{m}{n}}{\ln \omega} \right),$$

and $-\ln(2 - p) = \mathcal{O}\left(\ln \ln \frac{m}{n} - \ln \ln \omega\right) = \mathcal{O}\left(\ln \ln rm - \ln \ln \omega\right)$. □

Note that, for practical purposes, a good preconditioner has the approximation ratio in a constant factor away from the optimal $\sqrt{n}$. Note that Corollary 5.10 states that we can find such a matrix in $\mathcal{O}\left(\ln \frac{m}{n} \ln \ln(rm)\right)$ iterations of Algorithm 3.

### 5.3 Application to quartic problems

We now detail the consequences of the previous results on the preconditioning of quartics functions of the form (5.1).

**Improvement of $B_*$ over $B(\tau^{(0)})$.** Computing the optimal parameters $\tau^*$ requires running Algorithm 3 and might be costly. To evaluate the benefits, let us compare with the simpler *uniform* choice $\tau^{(0)} = (m^{-\frac{1}{2}}, \dots m^{-\frac{1}{2}})$, which corresponds to the inital point of the method. Writing $B^{(0)} = m^{-\frac{1}{2}} \sum_{i=1}^m B_i$ the corresponding preconditioner, Proposition 5.3 yields

$$\|x\|^2_{B^{(0)}} \leq \rho(x)^{1/2} \leq \left( \max_{i=1 \dots m} m \, \ell_i(\tau^{(0)}; \mathcal{B}) \right)^{1/2} \|x\|^2_{B^{(0)}} = \sqrt{m \gamma(\mathcal{B})} \, \|x\|^2_{B^{(0)}},$$

where we define $\gamma(\mathcal{B})$ to be the *coherence* of $\mathcal{B} = (B_1, \dots, B_m)$:

$$\gamma(\mathcal{B}) = \max_{i=1 \dots m} \ell_i(\tau^{(0)}; \mathcal{B}) = \max_{i=1 \dots m} \operatorname{Tr} \left[ \left( \sum_{i=1}^m B_i \right)^{-1} B_i \right], \tag{5.15}$$

The condition number of $\rho$ with respect to the norm induced by $B^{(0)}$ is therefore upper bounded by $\sqrt{m\gamma(\mathcal{B})}$. By Lemma 5.2 we have $\frac{n}{m} \leq \gamma(\mathcal{B}) \leq r$, therefore

$$\underbrace{\sqrt{n}}_{\text{condition number w.r.t. } B_*} \leq \underbrace{\sqrt{m\gamma(\mathcal{B})}}_{\text{condition number w.r.t. } B^{(0)}} \leq \sqrt{mr}.$$

Therefore, if the coherence of $\mathcal{B}$ is small, i.e. close to $n/m$, the approximation quality of $\tau^{(0)}$ and $\tau^*$ are similar. They become different if $\gamma(\mathcal{B})$ is high, and if $mr \gg n$. This is the situation where one could benefit from using $B(\tau^*)$ instead of $B^{(0)}$.

Note that if the operators $B_i$ are of rank 1 ($r = 1$) with $B_i = a_i a_i^T$, then $\gamma(\mathcal{B})$ coincides with the **matrix coherence** of $[a_1 \dots a_m]^T$, as defined in statistics and numerical Linear Algebra (Mahoney, 2011).

**Preconditioning for distance matrix completion.** Recall that for Problem (DMC), the quartic function writes for $X \in \mathbb{R}^{N\times r}$

$$\rho(X) = \sum_{(i,j)\in\Omega} \|x_i - x_j\|_2^4 + \frac{1}{N^2}\|\sum_{i=1}^N x_i\|_2^4,$$

where $x_1 \dots x_N$ are the rows of $X$ and $\Omega \subset \{1 \dots N\}^2$. We can write $\rho$ in the form (5.1) as $\rho(X) = \sum_{(i,j)\in\Omega}\langle B_{ij}(X), X\rangle^2 + \frac{1}{N}\langle \mathbf{1}_N\mathbf{1}_N^T X, X\rangle$ where the operators $B_{ij} : \mathbb{R}^{N\times r} \to \mathbb{R}^{N\times r}$ induce the quadratic forms $\langle B_{ij}(X), X\rangle = \|x_i - x_j\|_2^2$. The norm induced by the *uniform preconditioner* $B^{(0)}$ writes

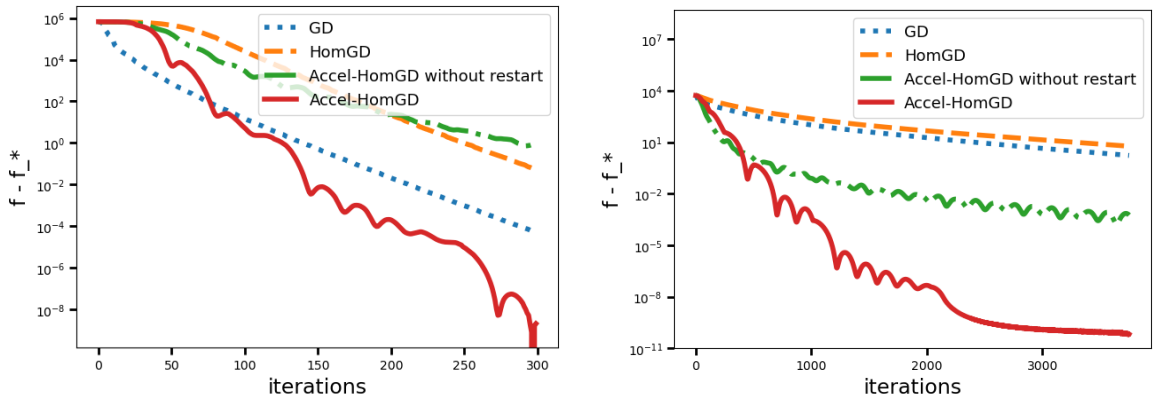
$$\begin{aligned}\|X\|_{B^{(0)}}^2 &= \frac{1}{\sqrt{|\Omega|+1}}\left(\sum_{(i,j)\in\Omega}\|x_i - x_j\|^2 + \|\sum_{i=1}^N x_i\|^2\right)\\ &= \frac{1}{\sqrt{|\Omega|+1}}\sum_{k=1}^r \langle\left(\mathcal{L}(\Omega) + \mathbf{1}_N\mathbf{1}_N^T\right)x_{\cdot k}, x_{\cdot k}\rangle,\end{aligned}$$

where $x_{\cdot k}$ is the $k$-th column of $X$ and $\mathcal{L}(\Omega) \in \mathbb{R}^{N\times N}$ is the Laplacian matrix of the graph with vertices $\{1 \dots N\}$ and edge set $\Omega$ Chung (1997). Then, one can show that the coherence of $\mathcal{B}$ as defined by (5.15) is equal to

$$\gamma(\mathcal{B}) = r\gamma(\Omega), \quad \gamma(\Omega) = \max_{(i,j)\in\Omega}\langle L(\Omega)^+(e_i - e_j), e_i - e_j\rangle,$$

where $\mathcal{L}(\Omega)^+$ denotes the pseudoinverse of the Laplacian matrix and $e_1 \dots e_N$ are the canonical basis vectors of $\mathbb{R}^N$. The quantity $\gamma(\Omega) \in [\frac{N}{|\Omega|}, 1]$ is called the graph *effective resistance* in spectral graph theory Ellens et al. (2011). $\gamma(\Omega)$ is equal to one if there exists an edge $(i,j)$ such that removing that edge from the graph would make it disconnected. On the contrary, $\gamma(\Omega)$ is low if all edges are of equal importance in determining the connectiveness of the graph.

According to the previous paragraph, the condition number of $\rho$ w.r.t. the norm induced by $B^{(0)}$ is $\kappa_0 = \sqrt{r|\Omega|\gamma(\Omega)}$, while the choice $B_*$ guarantees $\kappa^* = \sqrt{Nr}$. Therefore, using the optimal preconditioner $B^*$ is beneficial if the graph has **high effective resistance**, i.e. $|\Omega|\gamma(\Omega) \gg N$.




(a) $\frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} = 5$ (b) $\frac{\sigma_{\max}(A)}{\sigma_{\min}(A)} = 50$

Figure 1: Comparison of different methods on the synthetic quartic problem (6.1) for different condition numbers of $A$. We plot the objective gap $f(x_k) - f_*$, where $f_*$ is estimated as the minimal objective value computed by all methods. We observe that the restart scheme used in `Accel-HomGD` (Algorithm 2) is beneficial.

# 6 Numerical experiments

We consider the following quartic problem

$$\min_{x \in \mathbb{R}^n} \rho(x) - \langle c, x \rangle, \quad \rho(x) = \sum_{i=1}^{m} \langle a_i, x \rangle^4, \tag{6.1}$$

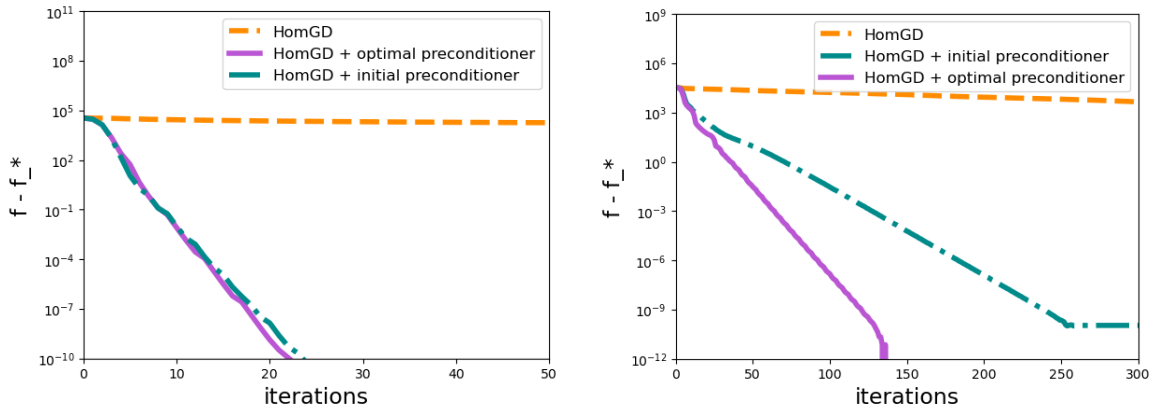
where $a_1 \dots a_m, c \in \mathbb{R}^n$ are generated randomly. We propose to examine the effect of homogenization, acceleration and preconditioning on different instances.

## 6.1 Homogenization and acceleration

We consider four different algorithms for solving (6.1). First, we implement our method `HomGD` (Algorithm 1) and its accelerated counterpart `Accel-HomGD` (Algorithm 2). We also try Algorithm 2 **without restart**, in order to evaluate the practical relevance of the restart technique, and to prove that it is not a mere theoretical construction. We compare these methods to `GD`, the standard Euclidean gradient descent method. For all variants, we use the usual Armijo line search strategy to adjust the step size.

We generate instances of (6.1) with random matrices $A \in \mathbb{R}^{m \times n}$ whose singular values are sampled uniformly in the interval $[\sigma_{\min}, 1]$, and random unit Gaussian vectors $c$. We choose $(m, n) = (2000, 1000)$. Figure 1 reports the results for two different values of $\sigma_{\min}$.

We observe (i) that `GD` and `HomGD` have similar practical performance, and therefore that the advantages of `HomGD` are mostly theoretical (convergence guarantees for fixed step size which does not depend on domain radius); (ii) that accelerated variants improve convergence speed for ill-conditioned problems; (iii) that the restart scheme used in Algorithm 2 has a significant impact on convergence speed, and therefore is not just a theoretical construct.




(a) low coherence: $\gamma(A) \approx \frac{n}{m}$

(b) high coherence: $\gamma(A) \approx 1$

Figure 2: Performance of the homogenized gradient method (Algorithm 1) on Problem (6.1) with three different preconditioners: `HomGD`uses the standard Euclidean norm ($B = I_n$), `HomGD + optimal preconditioner`uses the norm induced by $B^*$ as computed by Algorithm 3, while `HomGD + initial preconditioner`uses the norm induced by $B^{(0)}$, the initial iterate of Algorithm 3. As predicted by Section 5.3, the impact of using $B^*$ over $B^{(0)}$ is significant only for highly coherent matrices.

## 6.2 Preconditioning

Let us evaluate performance of the preconditioning scheme of Section 5. Recall that we compared in Section 5.3 two preconditioners:

$$\text{Initial preconditioner: } B^{(0)} = \frac{1}{\sqrt{m}} \sum_{i=1}^{m} a_i a_i^T, \qquad \kappa_0 = \sqrt{m\gamma(A)} \in [\sqrt{n}, \sqrt{mr}];$$

$$\text{Optimal preconditioner: } B^*, \qquad \kappa_* = \sqrt{n},$$

where $\gamma(A) = \max_{i=1\dots m} \langle (A^T A)^{-1} a_i, a_i \rangle$ is the *coherence* of matrix $A \in \mathbb{R}^{m \times n}$ with rows $a_1, \dots a_m$. Additionnally, since $\sigma_{\min}(A)^2 I_n \preceq B^{(0)} \preceq \sigma_{\max}(A)^2 I_n$, we can also estimate the condition number w.r.t. the standard Euclidean norm:

$$\text{No preconditioner } (B = I_n): \qquad \kappa_{I_n} \leq \frac{\sigma_{\max}(A)^2}{\sigma_{\min}(A)^2} \sqrt{m\gamma(A)}.$$

We run Algorithm 3 to compute $B^*$, and compare its performance to $B^{(0)}$ and $I_n$ when used with the homogenized gradient method (Algorithm 1).

We consider problem (6.1) with $m = 500$, $n = 20$. We generate random matrices $A$ which are ill-conditioned, with $\sigma_{\max}(A)/\sigma_{\min}(A) = 100$, and with different coherence values[3]. Figure 2 reports the results for a matrix with high coherence ($\gamma(A) \approx 1$) and a low-coherence matrix ($\gamma(A) \approx \frac{n}{m}$).

We observe that (i) since the spectral gap of matrix $A$ is large, using the norm induced by either $B_*$ or $B^{(0)}$ significantly improves performance over the standard Euclidean norm; (ii) the gap between $B_*$ and $B^{(0)}$ is significant only for highly-coherent matrices, as predicted by theory.

[3]In order to generate matrices $A$ with given coherence, we use the approach of (Ipsen and Wentworth, 2012, Section 6.1.)

# 7 Conclusion and perspectives

We have demonstrated that certain quartic convex problems could be efficiently solved by transforming them into a equivalent *homogenized* problem. The homogenized problem being smooth and convex, it can be solved with standard gradient methods. Our algorithms enjoy convergence guarantees with **fixed step size** and allow for **acceleration**. Previous methods, such as Euclidean or Bregman gradient descent, were not able satisfy both criteria simultaneously on quartic problems.

Note that we did not covert the case when the operator $Q$ is positive semidefinite, that is $\alpha = 0$. Under additional technical conditions, we could extend our method to handle this situation. However, we would not benefit from improved rates over those of standard gradient methods for general smooth convex functions.

Naturally, the next goal is to extend this technique to the class of general convex quartic polynomials of the form

$$\min_{x \in E} Q[x]^4 + P[x]^3 + A[x,x] + \langle c, x \rangle, \tag{7.1}$$

where $Q, P, A$ are multilinear maps of respective degrees 4, 3 and 2, as our current homogenization procedure cannot handle terms of degree 2 and 3 in the objective.

In our second contribution, we studied a method for finding a appropriate preconditioner $B_*$ for the quartic function by using a variant of the Lewis weights. One could ask if this technique can be extended to general polynomials of the form (7.1). Another possible direction for future research is to find a more efficient method for computing $\tau^*$, as the complexity of $\tilde{\mathcal{O}}(mrn^2)$ can be prohibitive for large-scale problems. A way to achieve this would be to compute the quantities $\ell_i(\tau; \mathcal{B})$ only approximatively.

**Acknowledgements.** The authors thank the two anonymous reviewers for their constructive comments which greatly contributed to improve the paper quality.

This paper has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 788368).

It was also supported by Multidisciplinary Institute in Artificial intelligence MIAI@Grenoble Alpes (ANR-19-P3IA-0003).

# References

Le Thi Hoai An and Pham Dinh Tao. The dc (difference of convex functions) programming and dca revisited with dc models of real world nonconvex optimization problems. *Annals of Operations Research*, 133:23–46, 2005.

K. M. Anstreicher. Ellipsoidal approximations of convex sets based on the volumetric barrier. *Mathematics of Operations Research*, 24:193–203, 1999.

Heinz H. Bauschke, Jérôme Bolte, and Marc Teboulle. A descent lemma beyond lipschitz gradient continuity: First-order methods revisited and applications. *Mathematics of Operations Research*, 42:330–348, 2017.

Amir Beck. *First-Order Methods in Optimization*. Society for Industrial and Applied Mathematics, 2017.

Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146:459–494, 2014.

Jérôme Bolte, Trong Phong Nguyen, Juan Peypouquet, and Bruce W. Suter. From error bounds to the complexity of first-order descent methods for convex functions. *Mathematical Programming*, 165:471–507, 2017.

Jérôme Bolte, Shoham Sabach, Marc Teboulle, and Yakov Vaisbourd. First order methods beyond convex-

ity and lipschitz gradient continuity with applications to quadratic inverse problems. *SIAM Journal on Optimization*, 28:2131–2151, 2018.

J. Bourgain, J. Lindenstrauss, and V. Milman. Approximation of zonoids by zonotopes. *Acta Mathematica*, 162(0):73–141, 1989.

Brian Bullins. Fast minimization of structured convex quartics. *arXiv preprint arXiv:1812.10349*, 2018.

Emmanuel J. Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61:1985–2007, 2015.

Fan RK Chung. *Spectral graph theory*, volume 92. American Mathematical Soc., 1997.

Michael B. Cohen and Richard Peng. $\ell_p$ row sampling by lewis weights. pages 183–192. ACM, 2015.

Nikita Doikov. Lower complexity bounds for minimizing regularized functions. *arXiv preprint arXiv:2202.04545v1*, 2022.

Ivan Dokmanic, Reza Parhizkar, Juri Ranieri, and Martin Vetterli. Euclidean distance matrices: Essential theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 32:12–30, 2015.

Radu-Alexandru Dragomir, Alexandre D'Aspremont, and Jérôme Bolte. Quartic first-order methods for low-rank minimization. *Journal of Optimization Theory and Applications*, 189:341–362, 2021a.

Radu-Alexandru Dragomir, Adrien Taylor, Alexandre D'Aspremont, and Jérôme Bolte. Optimal complexity and certification of bregman first-order methods. *Mathematical Programming*, 194:41–83, 2021b.

W. Ellens, F. M. Spieksma, P. Van Mieghem, A. Jamakovic, and R. E. Kooij. Effective graph resistance. *Linear Algebra and Its Applications*, 435:2491–2506, 2011.

Haw Ren Fang and Dianne P. O'Leary. Euclidean distance matrix completion problems. *Optimization Methods and Software*, 27:695–717, 2012.

Robert M. Freund and Haihao Lu. New computational guarantees for solving convex optimization problems with first order methods, via a function growth condition measure. *Mathematical Programming*, 170: 445–477, 2018.

Filip Hanzely, Peter Richtarik, and Lin Xiao. Accelerated bregman proximal gradient methods for relatively smooth convex optimization. *Computational Optimization and Applications*, 2021.

Zhaoshui He, Shengli Xie, Rafal Zdunek, Guoxu Zhou, and Andrzej Cichocki. Symmetric nonnegative matrix factorization: Algorithms and applications to probabilistic clustering. *IEEE Transactions on Neural Networks*, 22:2117–2131, 2011.

Hadrien Hendrikx, Lin Xiao, Sébastien Bubeck, Francis Bach, and Laurent Massoulié. Statistically preconditioned accelerated gradient method for distributed optimization. pages 4203–4227, 2020.

Anatoli Iouditski and Yuri Nesterov. Primal-dual subgradient methods for minimizing uniformly convex functions. *arXiv preprint arXiv:401.1792*, 2014.

Ilse C. F. Ipsen and Thomas Wentworth. The effect of coherence on sampling from matrices with orthonormal columns, and preconditioned least squares problems. 2012.

Fritz John. Extremum problems with inequalities as subsidiary conditions. *Extremum Problems with Inequalities as Side Conditions, Studies and Essays, Courant Anniversary Volume*, 1948.

Leonid G Khachiyan. Rounding of polytopes in the real number model of computation. *Mathematics of Operations Research*, 21:307–320, 1996.

Da Kuang, Sangwoon Yun, and Haesun Park. Symnmf: nonnegative low-rank approximation of a similarity matrix for graph clustering. *Journal of Global Optimization*, 62:545–574, 2015.

Daniel Ralph Lewis. Finite dimensional subspaces of $l_p$. *Studia Mathematica*, 63:207–212, 1978.

Haihao Lu, Robert M. Freund, and Yurii Nesterov. Relatively-smooth convex optimization by first-order methods, and applications. *SIAM Journal on Optimization*, 28:333–354, 2018.

Michael W. Mahoney. Randomized algorithms for matrices and data. *Foundations and Trends in Machine Learning*, 3:123–224, 2011.

Per Gunnar Martinsson and Joel A. Tropp. Randomized numerical linear algebra: Foundations and algorithms. *Acta Numerica*, 29:403–572, 2020.

Mahesh Chandra Mukkamala and Peter Ochs. Beyond alternating updates for matrix factorization with inertial bregman proximal gradient algorithms. 2019.

Arkadi Nemirovski and David B. Yudin. *Problem Complexity and Method Efficiency in Optimization*. Wiley–Blackwell, 1983.

Yurii Nesterov. A method of solving a convex programming problem with convergence rate $o(1/k^2)$. *Soviet Mathematics Doklady*, 1983.

Yurii Nesterov. Rounding of convex sets and efficient gradient methods for linear programming problems. *Optimization Methods and Software*, 23:109–128, 2008.

Yurii Nesterov. Superfast second-order methods for unconstrained convex optimization. *CORE Discussion Paper*, 2020.

Yurii Nesterov. Quartic regularity. *CORE Discussion Paper*, 2022.

Hou Duo Qi and Xiaoming Yuan. Computing the nearest euclidean distance matrix with low embedding dimensions. *Mathematical Programming*, 147:351–389, 2013.

Vincent Roulet and Alexandre d'Aspremont. Sharpness, restart and acceleration. Advances in Neural Information Processing Systems 30 (NIPS 2017), 2017.

Ohad Shamir, Nati Srebro, and Tong Zhang. Communication-efficient distributed optimization using an approximate Newton-type method. In *International Conference on Machine Learning*, 2014.

Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry N Chapman, and Jianwei Miao. Phase retrieval with application to optical imaging. *arXiv preprint arXiv:1402.7350v1*, pages 1–25, 2014.

Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, pages 1–68, 2017.

Michel Talagrand. Embedding subspaces of $l_p$ in $\ell_p^n$. In *Geometric Aspects of Functional Analysis*, pages 311–326. Birkhäuser Basel, 1995.

Pham Dinh Tao and El Bernoussi Souad. Algorithms for solving a class of nonconvex optimization problems. methods of subgradients. In *Fermat Days 85: Mathematics for Optimization*, pages 249–271. Elsevier, 1986.

Michael J. Todd. *Minimum-Volume Ellipsoids: Theory and Algorithms*. MOS-SIAM Series on Optimization, 2016.