A Greedy Algorithm for Subspace Approximation Problem

Nguyen Kim Thang
IBISC, Univ Evry, University Paris Saclay
Evry, France
thang@ibisc.fr
https://orcid.org/0000-0002-6085-9453

Abstract

In the subspace approximation problem, given m points in $\mathbb{R}^n$ and an integer $k \leq n$, the goal is to find a k-dimension subspace of $\mathbb{R}^n$ that minimizes the ℓ_p-norm of the Euclidean distances to the given points. This problem generalizes several subspace approximation problems and has applications from statistics, machine learning, signal processing to biology. Deshpande et al. [4] gave a randomized $O(\sqrt{p})$-approximation and this bound is proved to be tight assuming NP $\neq$ P by Guruswami et al. [7]. It is an intriguing question of determining the performance guarantee of deterministic algorithms for the problem. In this paper, we present a simple deterministic $O(\sqrt{p})$-approximation algorithm with also a simple analysis. That definitely settles the status of the problem in term of approximation up to a constant factor. Besides, the simplicity of the algorithm makes it practically appealing.

2012 ACM Subject Classification Theory of computation Approximation algorithms analysis

Keywords and phrases Approximation Algorithms, Subspace Approximation

Digital Object Identifier 10.4230/LIPIcs.SWAT.2018.30

Funding Research supported by the ANR project OATA n° ANR-15-CE40-0015-01.

1 Introduction

Massive data in high dimension emerge naturally in many domains from machine learning to biology. It has been observed that although data lie in high-dimensional spaces, in practice they have low intrinsic dimension. Dimension-reduction algorithms are essential in many domains such as image processing, personalized medicine, etc. In this paper, we consider the following subspace approximation problem in the context of capturing the underlying low-dimensional structures of given data.

Subspace Problem. Given points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$ and integers $p \geq 1$ and $0 \leq k \leq n$. Find a k-dimensional linear subspace W that minimizes the ℓ_p-norms of Euclidean distances of these points to W, i.e.,

$$\min_{W:\dim(W)=k} \left(\sum_{i=1}^{m} d(\boldsymbol{a}_i, W)^p \right)^{1/p}$$

The Subspace Problem, which is introduced by Deshpande et al. [4], is a generalization of several sub-space approximation problems which have been widely studied. For example, the well-known Least Square Fit Problem is a particular case. In the latter, given a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $0 \leq k \leq n$, find a matrix $\boldsymbol{B} \in \mathbb{R}^{m \times n}$ of rank at most k that minimizes the Frobenius norm of the difference $\|\boldsymbol{A} - \boldsymbol{B}\|_F := \left(\sum_{i,j}(A_{ij} - B_{ij})^2\right)^{1/2}$. Taking the rows of $\boldsymbol{A}$ to be $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m$ and $p = 2$, Subspace Problem reduces to Least Square Fit Problem. Another

16th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT 2018).
Editor: David Eppstein; Article No. 30; pp. 30:1–30:7
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

special case of Subspace Problem is Radii Problem. In the latter, given points $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m \in \mathbb{R}^n$, their *outer* $(n-k)$ *radius* is the minimum, over all k-dimensional linear subspaces, of the maximum Euclidean distances of these points to the subspace. This problem is equivalent to Subspace Problem with $p = \infty$. Moreover, Subspace Problem is related to other problems such as the L_p-Grothendieck Problem [8] and ℓ_p-Regression Problem [1, 5, 3]. We refer the reader to [4] for more details about the connection between these problems.

Deshpande et al. [4] introduced the Subspace Problem and gave a randomized $O(\sqrt{p})$-approximation algorithm. In their approach, they consider a convex relaxation that optimizes over positive semidefinite matrices and the rank constraint is replaced by a trace constraint. Subsequently, the solution $\boldsymbol{X}$ of the convex relaxation is rounded to a matrix of suitable rank. Intuitively, the authors divide singular vectors of the solution $\boldsymbol{X}$ into several bins and constructs one vector for each bin by taking a Bernoulli random linear combination of vectors within each bin. The analysis is carried out by powerful techniques coupled with properties of the p^{th}-moment of sums of Bernoulli random variables. Besides, Deshpande et al. [4] proved that the Subspace Problem is hard to approximate within a factor $\Omega(\sqrt{p})$ assuming the Unique Games Conjecture (UGC). Later on, bypassing the need for UGC, Guruswami et al. [7] showed that the problem is indeed NP-hard to approximate within a factor $\Omega(\sqrt{p})$.

Our Contribution. In this paper, we present a deterministic greedy algorithm with the same $O(\sqrt{p})$-approximation guarantee. Informally, at any step, the algorithm greedily extends the subspace in order to minimizes the marginal cost of the objective functions. The algorithm (and also the analysis) is extremely simple, which makes it practically appealing. Besides, our algorithm is deterministic whereas the one in [4] is randomized. The analysis is based on a smooth inequality (Lemma 2), which has been originally used in the context of algorithmic game theory in order to bound the quality of equilibrium (price of anarchy) in scheduling games [2]. This allows us to give a deterministic algorithm instead of the randomized one [4] which crucially relies on concentration inequalities in functional analysis in order to bound the moments of sums of Bernoulli random variables. Our result definitely settles the status of the problem in term of approximation up to a constant factor.

Related works. The most closely related to our paper is [4] where the results have been summarized earlier. For the Least Square Fit Problem, the optimal subspace is spanned by the top k right singular vector of $\boldsymbol{A}$ and that can be computed in time $O(\min\{n^2m, nm^2\})$ [6]. For the Radii Problem, $O(\sqrt{\log m})$-approximation with $k = n-1$ can be implied from the works of Nesterov [10] and Nemirovski et al. [9]. Later on, Varadarajan et al. [11] gave an $O(\sqrt{\log m})$-approximation algorithm for arbitrary k. Note that it is well-known that ℓ_∞-norm can be approximated by $\ell_{\log m}$-norm up to a constant factor. Hence, $O(\sqrt{\log m})$-approximation can be deduced from [4] (so our work) by choosing $p = \log m$.

2 Greedy Algorithm

As in [4], we use a formulation of the Subspace Problem in terms of the orthogonal complement of the subspace W. Specifically, let $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{n-k}$ be an orthonormal basis for the orthogonal complement. Let $\boldsymbol{Z} \in \mathbb{R}^{n \times (n-k)}$ be the matrix with the j^{th} column vector $\boldsymbol{z}_j$. Then $d(\boldsymbol{a}_i, W) = \|\boldsymbol{a}_i^T \boldsymbol{Z}\|_2$. Hence, the problem is to find an orthonormal basis $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{n-k}$ of a

$(n-k)$-dim vector space V so that the corresponding matrix $\boldsymbol{Z}$ minimizes

$$\sum_{i=1}^{m} \|\boldsymbol{a}_i^T \boldsymbol{Z}\|_2^p = \sum_{i=1}^{m} \left(\sum_{\ell=1}^{n-k} (\boldsymbol{a}_i^T \boldsymbol{z}_\ell)^2 \right)^{p/2}$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{n-k} \left[\left(\sum_{\ell=1}^{j} (\boldsymbol{a}_i^T \boldsymbol{z}_\ell)^2 \right)^{p/2} - \left(\sum_{\ell=1}^{j-1} (\boldsymbol{a}_i^T \boldsymbol{z}_\ell)^2 \right)^{p/2} \right],$$

where conventionally the sum with no term equals 0.

Algorithm. Initially, the subspace $U_0 = \emptyset$. For $1 \leq j \leq n-k$, choose a vector $\boldsymbol{u}_j \neq \boldsymbol{0}$ orthonormal to the subspace U_{j-1} spanned by $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_{j-1}$ such that it minimizes the marginal increase of the objective, i.e.,

$$\boldsymbol{u}_j \in \arg \min_{\boldsymbol{z} \perp U_{j-1}} \sum_{i=1}^{m} \left[\left((\boldsymbol{a}_i^T \boldsymbol{z})^2 + \sum_{\ell=1}^{j-1} (\boldsymbol{a}_i^T \boldsymbol{u}_\ell)^2 \right)^{p/2} - \left(\sum_{\ell=1}^{j-1} (\boldsymbol{a}_i^T \boldsymbol{u}_\ell)^2 \right)^{p/2} \right]$$

In fact, vector $\boldsymbol{u}_j$ can be computed by solving a convex program. Specifically, let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-j+1}\}$ be an arbitrary orthogonal basis of the vector space $U_{j-1}^{\perp}$ (which is orthogonal to U_{j-1}). Computing $\boldsymbol{u}_j$ is equivalent to computing the coefficients $b_1, \ldots, b_{n-j+1}$ in the decomposition of $\boldsymbol{u}_j$ in the basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-j+1}\}$. The convex program is

$$\min_{b_1,\ldots,b_{n-j+1}} \sum_{i=1}^{m} \left[\left(\left(\boldsymbol{a}_i^T \cdot \sum_{h=1}^{n-j+1} b_h \boldsymbol{e}_h \right)^2 + \sum_{\ell=1}^{j-1} (\boldsymbol{a}_i^T \boldsymbol{u}_\ell)^2 \right)^{p/2} - \left(\sum_{\ell=1}^{j-1} (\boldsymbol{a}_i^T \boldsymbol{u}_\ell)^2 \right)^{p/2} \right]$$

$$\sum_{h=1}^{n-j+1} b_h^2 = 1$$

$$b_1, \ldots, b_{n-j+1} \in \mathbb{R}$$

Note that in the convex program, variables are $b_1, \ldots, b_{n-j+1}$ (the vectors $\boldsymbol{u}_\ell$'s, $\boldsymbol{e}_h$'s have been already determined).

Analysis. Let $\boldsymbol{V}$ be an optimal $n \times (n-k)$ matrix and V be the corresponding vector space spanned by column vectors of $\boldsymbol{V}$. In the remaining, we will show that

$$\left(\sum_{i=1}^{m} \|\boldsymbol{a}_i^T \boldsymbol{U}\|_2^p \right)^{1/p} \leq O(\gamma_p) \cdot \left(\sum_{i=1}^{m} \|\boldsymbol{a}_i^T \boldsymbol{V}\|_2^p \right)^{1/p}$$

First, we recall the following standard lemma.

▶ **Lemma 1.** *For any $(n-k)$-dim subspace V, there exists an orthonormal basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-k}\}$ of V such that $\boldsymbol{v}_j$ is orthogonal to U_{j-1} for $1 \leq j \leq n-k$.*

Proof. We construct an orthogonal basis $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-k}\}$ of V by induction. The orthonormal basis is obtained by standard normalizing procedure. For $j = 1$, any arbitrary vector $\boldsymbol{v}_1 \in V$ is perpendicular to U_0. Assume that vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j$ for $j < (n-k)$ have been constructed so that they satisfy the lemma. Since the subspace V has dimension $(n-k)$ which is strictly larger than j, there exits a vector $\boldsymbol{w}_{j+1} \in V$ which is independent to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j$. Define vector $\boldsymbol{v}_{j+1} := \boldsymbol{w}_{j+1} - \mathsf{Pr}_{\boldsymbol{u}_j}(\boldsymbol{w}_{j+1})$ where $\mathsf{Pr}_{\boldsymbol{u}_j}(\boldsymbol{w}_{j+1})$ is the projection of vector $\boldsymbol{w}_{j+1}$ onto the subspace U_j. So $\boldsymbol{v}_{j+1}$ is orthogonal to $\boldsymbol{u}_j$ and is independent to $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_j$. ◀

▶ **Lemma 2.** *For any given vector $\boldsymbol{a}$, for arbitrary vectors $\boldsymbol{u}_\ell$ and $\boldsymbol{v}_\ell$ with $1 \leq \ell \leq n-k$, it holds that*

$$\sum_{j=1}^{n-k} \left[\left((a^T \boldsymbol{v}_j)^2 + \sum_{\ell=1}^{j-1} \left(a^T \boldsymbol{u}_\ell\right)^2 \right)^{p/2} - \left(\sum_{\ell=1}^{j-1} \left(a^T \boldsymbol{u}_\ell\right)^2 \right)^{p/2} \right] \leq \mu \left(\sum_{\ell=1}^{n-k} \left(a^T \boldsymbol{u}_\ell\right)^2 \right)^{p/2} + \lambda \left(\sum_{\ell=1}^{n-k} \left(a^T \boldsymbol{v}_\ell\right)^2 \right)^{p/2}$$

where $\lambda = O\big((\alpha \cdot \frac{p}{2})^{\frac{p}{2}-1}\big)$ for some constant α and $\mu = \frac{p/2-1}{p/2}$.

Proof. Denote $b_j = (a^T \boldsymbol{v}_j)^2$ and $c_j = (a^T \boldsymbol{u}_j)^2$ for $1 \leq j \leq n-k$. The lemma inequality reads

$$\sum_{j=1}^{n-k} \left[\left(b_j + \sum_{\ell=1}^{j-1} c_\ell \right)^{p/2} - \left(\sum_{\ell=1}^{j-1} c_\ell \right)^{p/2} \right] \leq \mu \left(\sum_{\ell=1}^{n-k} c_\ell \right)^{p/2} + \lambda \left(\sum_{\ell=1}^{n-k} b_\ell \right)^{p/2}$$

The inequality holds for $\lambda = O\big((\alpha \cdot \frac{p}{2})^{\frac{p}{2}-1}\big)$ for some constant α and $\mu = \frac{p/2-1}{p/2}$, which has been proved in [2] in the context of algorithmic game theory. For completeness, we put the proof of the above inequality in the appendix (Lemma 5). ◀

▶ **Theorem 3.** *The greedy algorithm is $O(\sqrt{p})$-approximation.*

Proof. Recall that $\boldsymbol{U}$ be the solution of the algorithm where the j^{th} column vector is $\boldsymbol{u}_j$ for $1 \leq j \leq n-k$. Let $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{n-k}$ be an orthonormal basis of V that satisfies Lemma 1. We have

$$\begin{aligned}
\sum_{i=1}^{m} \|\boldsymbol{a}_i^T \boldsymbol{U}\|_2^p &= \sum_{i=1}^{m} \sum_{j=1}^{n-k} \left[\left(\sum_{\ell=1}^{j} \left(\boldsymbol{a}_i^T \boldsymbol{u}_\ell\right)^2 \right)^{p/2} - \left(\sum_{\ell=1}^{j-1} \left(\boldsymbol{a}_i^T \boldsymbol{u}_\ell\right)^2 \right)^{p/2} \right] \\
&\leq \sum_{i=1}^{m} \sum_{j=1}^{n-k} \left[\left((\boldsymbol{a}_i^T \boldsymbol{v}_j)^2 + \sum_{\ell=1}^{j-1} \left(\boldsymbol{a}_i^T \boldsymbol{u}_\ell\right)^2 \right)^{p/2} - \left(\sum_{\ell=1}^{j-1} \left(\boldsymbol{a}_i^T \boldsymbol{u}_\ell\right)^2 \right)^{p/2} \right] \\
&\leq \sum_{i=1}^{m} \left[\mu \left(\sum_{\ell=1}^{n-k} \left(\boldsymbol{a}_i^T \boldsymbol{u}_\ell\right)^2 \right)^{p/2} + \lambda \left(\sum_{\ell=1}^{n-k} \left(\boldsymbol{a}_i^T \boldsymbol{v}_\ell\right)^2 \right)^{p/2} \right] \\
&= \mu \sum_{i=1}^{m} \|\boldsymbol{a}_i^T \boldsymbol{U}\|_2^p + \lambda \sum_{i=1}^{m} \|\boldsymbol{a}_i^T \boldsymbol{V}\|_2^p
\end{aligned}$$

The first inequality is due to the choice of the algorithm at any step j (note that $\boldsymbol{v}_j \perp U_{j-1}$ so $\boldsymbol{v}_j$ is a candidate at step j). The second inequality holds by Lemma 2 where $\lambda = O\big((\alpha \cdot \frac{p}{2})^{\frac{p}{2}-1}\big)$ for some constant α and $\mu = \frac{p/2-1}{p/2}$. Rearranging the terms and taking the p^{th}-root, we get

$$\left(\sum_{i=1}^{m} \|\boldsymbol{a}_i^T \boldsymbol{U}\|_2^p \right)^{1/p} \leq O(\sqrt{p}) \cdot \left(\sum_{i=1}^{m} \|\boldsymbol{a}_i^T \boldsymbol{V}\|_2^p \right)^{1/p}$$

◀

References

1 Kenneth L Clarkson. Subgradient and sampling algorithms for ℓ_1 regression. In *Proc. 16th Symposium on Discrete algorithms*, pages 257–266, 2005.

2 Johanne Cohen, Christoph Dürr, and Nguyen Kim Thang. Smooth inequalities and equilibrium inefficiency in scheduling games. In *Internet and Network Economics*, pages 350–363, 2012.

3 Anirban Dasgupta, Petros Drineas, Boulos Harb, Ravi Kumar, and Michael W Mahoney. Sampling algorithms and coresets for ℓ_p regression. *SIAM Journal on Computing*, 38(5):2060–2078, 2009.

4 Amit Deshpande, Madhur Tulsiani, and Nisheeth K Vishnoi. Algorithms and hardness for subspace approximation. In *Proc. 22nd Symposium on Discrete Algorithms*, pages 482–496, 2011.

5 Petros Drineas, Michael W Mahoney, and S Muthukrishnan. Sampling algorithms for ℓ_2 regression and applications. In *Proc. 17th Symposium on Discrete algorithm*, pages 1127–1136, 2006.

6 Gene H Golub and Charles F Van Loan. *Matrix computations*. Johns Hopkins University Press, 2012.

7 Venkatesan Guruswami, Prasad Raghavendra, Rishi Saket, and Yi Wu. Bypassing UGC from some optimal geometric inapproximability results. *ACM Transactions on Algorithms*, 12(1):6, 2016.

8 Guy Kindler, Assaf Naor, and Gideon Schechtman. The UGC hardness threshold of the L_p grothendieck problem. *Mathematics of Operations Research*, 35(2):267–283, 2010.

9 Arkadi Nemirovski, Cornelis Roos, and Tamás Terlaky. On maximization of quadratic form over intersection of ellipsoids with common center. *Mathematical Programming*, 86(3):463–473, 1999.

10 Yuri Nesterov. *Global quadratic optimization via conic relaxation*. Université Catholique de Louvain. Center for Operations Research and Econometrics [CORE], 1998.

11 Kasturi Varadarajan, Srinivasan Venkatesh, Yinyu Ye, and Jiawei Zhang. Approximating the radii of point sets. *SIAM Journal on Computing*, 36(6):1764–1776, 2007.

Appendix: Technical Lemmas

In this section, we show technical lemmas. The following lemma has been proved in [2]. We give it here for completeness.

▶ **Lemma 4** ([2]). *Let k be a positive integer. Let $0 < a(k) \leq 1$ be a function on k. Then, for any $x, y > 0$, it holds that*

$$y(x+y)^k \leq \frac{k}{k+1} a(k) x^{k+1} + b(k) y^{k+1}$$

where α is some constant and

$$b(k) = \begin{cases} \Theta\left(\alpha^k \cdot \left(\frac{k}{\log ka(k)}\right)^{k-1}\right) & \text{if } \lim_{k\to\infty}(k-1)a(k) = \infty, \quad (1a) \\ \Theta\left(\alpha^k \cdot k^{k-1}\right) & \text{if } (k-1)a(k) \text{ are bounded } \forall k, \quad (1b) \\ \Theta\left(\alpha^k \cdot \frac{1}{ka(k)^k}\right) & \text{if } \lim_{k\to\infty}(k-1)a(k) = 0. \quad (1c) \end{cases}$$

Proof. Let $f(z) := \frac{k}{k+1} a(k) z^{k+1} - (1+z)^k + b(k)$. To show the claim, it is equivalent to prove that $f(z) \geq 0$ for all $z > 0$.

We have $f'(z) = ka(k)z^k - k(1+z)^{k-1}$. We claim that the equation $f'(z) = 0$ has an unique positive root z_0. Consider the equation $f'(z) = 0$ for $z > 0$. It is equivalent to

$$\left(\frac{1}{z}+1\right)^k \cdot \frac{1}{z} = a(k)$$

The left-hand side is a strictly decreasing function and the limits when z tends to 0 and ∞ are ∞ and 0, respectively. As $a(k)$ is a positive constant, there exists an unique root $z_0 > 0$.

Observe that function f is decreasing in $(0, z_0)$ and increasing in $(z_0, +\infty)$, so $f(z) \geq f(z_0)$ for all $z > 0$. Hence, by choosing

$$b(k) = \left|\frac{k}{k+1} a(k) z_0^{k+1} - (1+z_0)^k\right| = (1+z_0)^{k-1}\left(1 + \frac{z_0}{k+1}\right) \quad (2)$$

it follows that $f(z) \geq 0 \ \forall z > 0$.

We study the positive root z_0 of equation

$$a(k)z^k - (1+z)^{k-1} = 0 \quad (3)$$

Note that $f'(1) = k(a(k) - 2^{k-1}) < 0$ since $0 < a(k) \leq 1$. Thus, $z_0 > 1$. For the sake of simplicity, we define the function $g(k)$ such that $z_0 = \frac{k-1}{g(k)}$ where $0 < g(k) < k-1$. Equation (3) is equivalent to

$$\left(1 + \frac{g(k)}{k-1}\right)^{k-1} g(k) = (k-1)a(k)$$

Note that $e^{w/2} < 1 + w < e^w$ for $w \in (0,1)$. For $w := \frac{g(k)}{k-1}$, we obtain the following upper and lower bounds for the term $(k-1)a(k)$:

$$e^{g(k)/2} g(k) < (k-1)a(k) < e^{g(k)} g(k) \quad (4)$$

Recall the definition of *Lambert W function*. For each $y \in \mathbb{R}^+$, $W(y)$ is defined to be solution of the equation $xe^x = y$. Note that, xe^x is increasing with respect to x, hence $W(\cdot)$ is increasing.

By definition of the Lambert W function and Equation (4), we get that

$$W\left((k-1)a(k)\right) < g(k) < 2W\left(\frac{(k-1)a(k)}{2}\right) \tag{5}$$

First, consider the case where $\lim_{k\to\infty}(k-1)a(k) = \infty$. The asymptotic sequence for $W(x)$ as $x \to +\infty$ is the following: $W(x) = \ln x - \ln\ln x + \frac{\ln\ln x}{\ln x} + O\left(\left(\frac{\ln\ln x}{\ln x}\right)^2\right)$. So, for large enough k, $W((k-1)a(k)) = \Theta(\log((k-1)a(k)))$. Since $z_0 = \frac{k-1}{g(k)}$, from Equation (5), we get $z_0 = \Theta\left(\frac{k}{\log(ka(k))}\right)$. Therefore, by (2) we have $b(k) = \Theta\left(\alpha^k \cdot \left(\frac{k}{\log ka(k)}\right)^{k-1}\right)$ for some constant α.

Second, consider the case where $(k-1)a(k)$ is bounded by some constants. So by (5), we have $g(k) = \Theta(1)$. Therefore $z_0 = \Theta(k)$ which again implies $b(k) = \Theta\left(\alpha^k \cdot k^{k-1}\right)$ for some constant α.

Third, we consider the case where $\lim_{k\to\infty}(k-1)a(k) = 0$. We focus on the Taylor series W_0 of W around 0. It can be found using the Lagrange inversion and is given by

$$W_0(x) = \sum_{i=1}^{\infty} \frac{(-i)^{i-1}}{i!} x^i = x - x^2 + O(1)x^3.$$

Thus, for k large enough $g(k) = \Theta((k-1)a(k))$. Hence, $z_0 = \Theta(1/a(k))$. Once again that implies $b(k) = \Theta\left(\alpha^k \cdot \frac{1}{ka(k)^k}\right)$ for some constant α. ◀

▶ **Lemma 5.** *For any sequences of non-negative real numbers $\{a_1, a_2, \ldots, a_n\}$ and $\{b_1, b_2, \ldots, b_n\}$ and for any polynomial g of degree k with non-negative coefficients, it holds that*

$$\sum_{i=1}^{n}\left[g\left(b_i + \sum_{j=1}^{i-1} a_j\right) - g\left(\sum_{j=1}^{i-1} a_j\right)\right] \leq \lambda(k)\cdot g\left(\sum_{i=1}^{n} b_i\right) + \mu(k)\cdot g\left(\sum_{i=1}^{n} a_i\right)$$

where $\mu(k) = \frac{k-1}{k}$ and $\lambda(k) = \Theta\left(k^{k-1}\right)$. The same inequality holds for $\mu(k) = \frac{k-1}{k\ln k}$ and $\lambda(k) = \Theta\left((\alpha\cdot k\ln k)^{k-1}\right)$ for some constant α.

Proof. Let $g(z) = g_0 z^k + g_1 z^{k-1} + \cdot + g_k$ with $g_t \geq 0\ \forall t$. The lemma holds since it holds for every z^t for $0 \leq t \leq k$. Specifically,

$$\begin{aligned}
\sum_{i=1}^{n}\left[g\left(b_i + \sum_{j=1}^{i-1} a_j\right) - g\left(\sum_{j=1}^{i-1} a_j\right)\right] &= \sum_{t=1}^{k} g_{k-t}\cdot\sum_{i=1}^{n}\left[\left(b_i + \sum_{j=1}^{i-1} a_j\right)^t - \left(\sum_{j=1}^{i-1} a_j\right)^t\right] \\
&\leq \sum_{t=1}^{k} g_{k-t}\cdot\left[t\cdot b_i\cdot\left(b_i + \sum_{j=1}^{i-1} a_j\right)^{t-1}\right] \leq \sum_{t=1}^{k} g_{k-t}\cdot\left[\lambda(t)\left(\sum_{i=1}^{n} b_i\right)^t + \mu(t)\left(\sum_{i=1}^{n} a_i\right)^t\right] \\
&\leq \lambda(k)\cdot g\left(\sum_{i=1}^{n} b_i\right) + \mu(k)\cdot g\left(\sum_{i=1}^{n} a_i\right)
\end{aligned}$$

The first inequality follows the convex inequality $(x+y)^{k+1} - x^{k+1} \leq (k+1)y(x+y)^k$. The second inequality follows Lemma 4 (Case (1b) and $a(k) = 1/(k+1)$). The last inequality holds since $\mu(t) \leq \mu(k)$ and $\lambda(t) \leq \lambda(k)$ for $t \leq k$. ◀