Statistical analysis of k-nearest neighbor collaborative recommendation

Gérard Biau, Benoît Cadre, Laurent Rouvière

▶ **To cite this version:**

Gérard Biau, Benoît Cadre, Laurent Rouvière. Statistical analysis of k-nearest neighbor collaborative recommendation. Annals of Statistics, 2010, 38 (3), pp.1568-1592. 10.1214/09-AOS759 . hal-00367480v2

HAL Id: hal-00367480
https://hal.science/hal-00367480v2

Submitted on 13 Oct 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical Analysis of k-Nearest Neighbor Collaborative Recommendation

Gérard BIAU [a,*], Benoît CADRE [b] and Laurent ROUVIÈRE [c]

[a] LSTA & LPMA
Université Pierre et Marie Curie – Paris VI
Boîte 158, 175 rue du Chevaleret
75013 Paris, France
gerard.biau@upmc.fr

[b] IRMAR, ENS Cachan Bretagne, CNRS, UEB
Campus de Ker Lann
Avenue Robert Schuman
35170 Bruz, France
Benoit.Cadre@bretagne.ens-cachan.fr

[c] CREST-ENSAI, IRMAR, UEB
Campus de Ker Lann
Rue Blaise Pascal - BP 37203
35172 Bruz Cedex, France
laurent.rouviere@ensai.fr

Abstract

Collaborative recommendation is an information-filtering technique that attempts to present information items that are likely of interest to an Internet user. Traditionally, collaborative systems deal with situations with two types of variables, users and items. In its most common form, the problem is framed as trying to estimate ratings for items that have not yet been consumed by a user. Despite wide-ranging literature, little is known about the statistical properties of recommendation systems. In fact, no clear probabilistic model even exists which would allow us to precisely describe the mathematical forces driving collaborative filtering. To provide an initial contribution to this, we propose to set out a general sequential stochastic model for collaborative recommendation. We offer an in-depth analysis of the so-called cosine-type nearest neighbor collaborative method, which is one of the most widely used algorithms in collaborative filtering, and

*Corresponding author.

analyze its asymptotic performance as the number of users grows. We establish consistency of the procedure under mild assumptions on the model. Rates of convergence and examples are also provided.

Index Terms — Collaborative recommendation – cosine-type similarity – nearest neighbor estimate – consistency – rate of convergence.

AMS 2000 Classification: 62G05, 62G20.

1 Introduction

Collaborative recommendation is a Web information-filtering technique that typically gathers information about your personal interests and compares your profile to other users with similar tastes. The goal of this system is to give personalized recommendations, whether this be movies you might enjoy, books you should read or the next restaurant you should go to.

There has been much work done in this area over the past decade since the appearance of the first papers on the subject in the mid-90's (Resnick et al. [13], Hill et al. [11], Shardanand and Maes [16]). Stimulated by an abundance of practical applications, most of the research activity to date has focused on elaborating various heuristics and practical methods (Breese et al. [4], Heckerman et al. [10], Salakhutdinov et al. [14]) so as to provide personalized recommendations and help Web users deal with information overload. Examples of such applications include recommending books, people, restaurants, movies, CDs and news. Websites such as amazon.com, match.com, movielens.org and allmusic.com already have recommendation systems in operation. We refer the reader to the surveys by Adomavicius and Tuzhilin [3] and Adomavicius et al. [2] for a broader picture of the field, an overview of results and many related references.

Traditionally, collaborative systems deal with situations with two types of variables, *users* and *items*. In its most common form, the problem is framed as trying to estimate *ratings* for items that have *not* yet been consumed by a user. The recommendation process typically starts by asking users a series of questions about items they liked or did not like. For example, in a movie recommendation system, users initially rate some subset of films they have already seen. Personal ratings are then collected in a matrix, where each row represents a user, each column an item, and entries in the matrix represent a given user's rating of a given item. An example is presented in Table 1, where ratings are specified on a scale from 1 to 10, and "NA" means that the user has not rated the corresponding film.

	Armageddon	Platoon	Rambo	Rio Bravo	Star wars	Titanic
Jim	NA	6	7	8	9	NA
James	3	NA	10	NA	5	7
Steve	7	NA	1	NA	6	NA
Mary	NA	7	1	NA	5	6
John	NA	7	NA	NA	3	1
Lucy	3	10	2	7	NA	4
Stan	NA	7	NA	NA	1	NA
Johanna	4	5	NA	8	3	9
Bob	NA	3	3	4	5	?

Table 1: A (subset of a) ratings matrix for a movie recommendation system. Ratings are specified on a scale from 1 to 10, and "NA" means that the user has not rated the corresponding film.

Based on this prior information, the recommendation engine must be able to automatically furnish ratings of as-yet unrated items and then suggest appropriate recommendations based on these predictions. To do this, a number of practical methods have been proposed, including machine learning-oriented techniques (e.g., Abernethy et al. [1]), statistical approaches (e.g., Sarwar et al. [15]) and numerous other ad hoc rules (Adomavicius and Tuzhilin [2]). The collaborative filtering issue may be viewed as a special instance of the problem of inferring the many missing entries of a data matrix. This field, which has very recently emerged, is known as the matrix completion problem, and comes up in many areas of science and engineering, including collaborative filtering, machine learning, control, remote sensing and computer vision. We will not pursue this promising approach, and refer the reader to Candès and Recht [6] and Candès and Plan [5] who survey the literature on matrix completion. These authors show in particular that under suitable conditions, one can recover an unknown low rank matrix from a nearly minimal set of entries by solving a simple convex optimization problem.

In most of the approaches, the crux is to identify users whose tastes/ratings are "similar" to the user we would like to advise. The similarity measure assessing proximity between users may vary depending on the type of application, but is typically based on a correlation or cosine-type approach (Sarwar et al. [15]).

Despite wide-ranging literature, very little is known about the statistical properties of recommendation systems. In fact, no clear probabilistic model

even exists allowing us to precisely describe the mathematical forces driving collaborative filtering. To provide an initial contribution to this, we propose in the present paper to set out a general stochastic model for collaborative recommendation and analyze its asymptotic performance as the number of users grows.

The document is organized as follows. In section 2, we provide a sequential stochastic model for collaborative recommendation and describe the statistical problem. In the model we analyze, unrated items are estimated by averaging ratings of users who are "similar" to the user we would like to advise. The similarity is assessed by a cosine-type measure, and unrated items are estimated using a k_n-nearest neighbor-type regression estimate, which is indeed one of the most widely used procedures in collaborative filtering. It turns out that the choice of the cosine proximity as a similarity measure imposes constraints on the model, which are discussed in section 3. Under mild assumptions, consistency of the estimation procedure is established in section 4, whereas rates of convergence are discussed in section 5. Illustrative examples are given throughout the document, and proofs of some technical results are postponed to section 6.

2 A model for collaborative recommendation

2.1 Ratings matrix and new users

Suppose that there are $d+1$ ($d \geq 1$) possible items, n users in the ratings matrix (i.e., the database) and that users' ratings take values in the set $(\{0\} \cup [1, s])^{d+1}$. Here, s is a real number greater than 1 corresponding to the maximal rating and, by convention, the symbol 0 means that the user has not rated the item (same as "NA"). Thus, the ratings matrix has n rows, $d+1$ columns and entries from $\{0\} \cup [1, s]$. For example, $n = 8$, $d = 5$ and $s = 10$ in Table 1, which will be our toy example throughout this section. Then, a new user Bob reveals some of his preferences for the first time, rating some of the first d items but *not* the $(d+1)$th (the movie Titanic in Table 1). We want to design a strategy to predict Bob's rating of Titanic using: (i) Bob's ratings of some (or all) of the other d movies and (ii) the ratings matrix. This is illustrated in Table 1, where Bob has rated 4 out of the 5 movies.

The first step in our approach is to model the preferences of new user Bob by a random vector $(\mathbf{X}, Y)$ of size $d+1$ taking values in the set $[1, s]^d \times [1, s]$. Within this framework, the random variable $\mathbf{X} = (X_1, \ldots, X_d)$ represents

Bob's preferences pertaining to the first d movies, whereas Y, the (unobserved) variable of interest, refers to the movie Titanic. In fact, as Bob does not necessarily reveals all his preferences at once, we do not observe the variable $\mathbf{X}$, but instead some "masked" version of it denoted hereafter by $\mathbf{X}^\star$. The random variable $\mathbf{X}^\star = (X_1^\star, \ldots, X_d^\star)$ is naturally defined by

$$X_j^\star = \begin{cases} X_j & \text{if } j \in M \\ 0 & \text{otherwise,} \end{cases}$$

where M stands for some non-empty random subset of $\{1, \ldots, d\}$ indexing the movies which have been rated by Bob. Observe that the random variable $\mathbf{X}^\star$ takes values in $(\{0\} \cup [1, s])^d$ and that $\|\mathbf{X}^\star\| \geq 1$, where $\|.\|$ denotes the usual Euclidean norm on $\mathbb{R}^d$. In the example of Table 1, $M = \{2, 3, 4, 5\}$ and (the realization of) $\mathbf{X}^\star$ is $(0, 3, 3, 4, 5)$.

We follow the same approach to model preferences of users already in the database (Jim, James, Steve, Mary, etc. in Table 1), who will therefore be represented by a sequence of independent $[1, s]^d \times [1, s]$-valued random pairs $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ from the distribution $(\mathbf{X}, Y)$. A first idea for dealing with potential non-responses of a user i in the ratings matrix ($i = 1, \ldots, n$) is to consider in place of $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})$ its masked version $\widetilde{\mathbf{X}}_i = (\widetilde{X}_{i1}, \ldots, \widetilde{X}_{id})$ defined by

$$\widetilde{X}_{ij} = \begin{cases} X_{ij} & \text{if } j \in M_i \cap M \\ 0 & \text{otherwise,} \end{cases} \tag{2.1}$$

where each M_i is the random subset of $\{1, \ldots, d\}$ indexing the movies which have been rated by user i. In other words, we only keep in $\mathbf{X}_i$ items corated by both user i *and* the new user — items which have not been rated by $\mathbf{X}$ and $\mathbf{X}_i$ are declared non-informative and simply thrown away.

However, this model, which is static in nature, does not allow to take into account the fact that, as time goes by, each user in the database may reveal more and more preferences. This will for instance typically be the case in the movie recommendation system of Table 1, where regular customers will update their ratings each time they have seen a new movie. Consequently, model (2.1) is not fully satisfying and must therefore be slightly modified to better capture the sequential evolution of ratings.

2.2 A sequential model

A possible dynamical approach for collaborative recommendation is based on the following protocol: users enter the database one after the other and

update their list of ratings sequentially in time. More precisely, we suppose that at each time $i = 1, 2, \ldots$, a new user enters the process and reveals his preferences for the first time, while the $i-1$ previous users are allowed to rate new items. Thus, at time 1, there is only one user in the database (Jim in Table 1), and the (non-empty) subset of items he decides to rate is modeled by a random variable M_1^1 taking values in $\mathcal{P}^\star(\{1, \ldots, d\})$, the set of non-empty subsets of $\{1, \ldots, d\}$. At time 2, a new user (James) enters the game and reveals his preferences according to a $\mathcal{P}^\star(\{1, \ldots, d\})$-valued random variable M_2^1, with the same distribution as M_1^1. At the same time, Jim (user 1) may update his list of preferences, modeled by a random variable M_1^2 satisfying $M_1^1 \subset M_1^2$. The latter requirement just means that the user is allowed to rate new items but not to remove his past ratings. At time 3, a new user (Steve) rates items according to a random variable M_3^1 distributed as M_1^1, while user 2 updates his preferences according to M_2^2 (distributed as M_1^2) and user 1 updates his own according to M_1^3, and so on. This sequential mechanism is summarized in Table 2.

	Time 1	Time 2	$\ldots$	Time i	$\ldots$	Time n
User 1	M_1^1	M_1^2	$\ldots$	M_1^i	$\ldots$	M_1^n
User 2		M_2^1	$\ldots$	M_2^{i-1}	$\ldots$	M_2^{n-1}
$\vdots$			$\ddots$	$\vdots$	$\vdots$	$\vdots$
User i				M_i^1	$\ldots$	M_i^{n+1-i}
$\vdots$					$\ddots$	$\vdots$
User n						M_n^1

Table 2: A sequential model for preference updating.

By repeating this procedure, we end up at time n with an upper triangular array $(M_i^j)_{1 \leq i \leq n, 1 \leq j \leq n+1-i}$ of random variables. A row in this array consists of a collection M_i^j of random variables for a given value of i, taking values in $\mathcal{P}^\star(\{1, \ldots, d\})$ and satisfying the constraint $M_i^j \subset M_i^{j+1}$. For a fixed i, the sequence $M_i^1 \subset M_i^2 \subset \ldots$ describes the (random) way user i sequentially reveals his preferences over time. Observe that the later inclusions are not necessarily strict, so that a single user is not forced to rate one more item at every single step.

Throughout the paper, we will assume that, for each i, the distribution of the sequence of random variables $(M_i^n)_{n \geq 1}$ is independent of i, and is therefore distributed as a generic random sequence denoted $(M^n)_{n \geq 1}$, satisfying $M^1 \neq$

$\emptyset$ and $M^n \subset M^{n+1}$ for all $n \geq 1$. For the sake of coherence, we assume that M^1 and M (see (2.1)) have the same distribution, i.e., the new abstract user $\mathbf{X}^\star$ may be regarded as a user entering the database for the first time. We will also suppose that there exists a positive random integer n_0 such that $M^{n_0} = \{1, \ldots, d\}$ and, consequently, $M^n = \{1, \ldots, d\}$ for all $n \geq n_0$. This requirement means that each user rates all d items after a (random) period of time. Last, we will assume that the pairs $(\mathbf{X}_i, Y_i)$, $i = 1, \ldots, n$, the sequences $(M_1^n)_{n\geq 1}$, $(M_2^n)_{n\geq 1}, \ldots$ and the random variable M are mutually independent. We note that this implies that the users' ratings are independent.

With this sequential point of view, improving on (2.1), we let the masked version $\mathbf{X}_i^{(n)} = (X_{i1}^{(n)}, \ldots, X_{id}^{(n)})$ of $\mathbf{X}_i$ be defined as

$$X_{ij}^{(n)} = \begin{cases} X_{ij} & \text{if } j \in M_i^{n+1-i} \cap M \\ 0 & \text{otherwise.} \end{cases}$$

Again, it is worth pointing out that, in the definition of $\mathbf{X}_i^{(n)}$, items which have not been corated by both $\mathbf{X}$ *and* $\mathbf{X}_i$ are deleted. This implies in particular that $\mathbf{X}_i^{(n)}$ may be equal to $\mathbf{0}$, the d-dimensional null vector (whereas $\|\mathbf{X}^\star\| \geq 1$ by construction).

Finally, in order to deal with possible non-answers of database users regarding the variable of interest (Titanic in our movie example), we introduce $(\mathcal{R}_n)_{n\geq 1}$, a sequence of random variables taking values in $\mathcal{P}^\star(\{1, \ldots, n\})$, such that $\mathcal{R}_n$ is independent of M and the sequences $(M_i^n)_{n\geq 1}$, and satisfying $\mathcal{R}_n \subset \mathcal{R}_{n+1}$ for all $n \geq 1$. In this formalism, $\mathcal{R}_n$ represents the subset, which is assumed to be non-empty, of users who have already provided information about Titanic at time n. For example, in Table 1, only James, Mary, John, Lucy and Johanna have rated Titanic and therefore (the realization of) $\mathcal{R}_n$ is $\{2, 4, 5, 6, 8\}$.

2.3 The statistical problem

To summarize the model so far, we have at hand at time n a sample of random pairs $(\mathbf{X}_1^{(n)}, Y_1), \ldots, (\mathbf{X}_n^{(n)}, Y_n)$ and our mission is to predict the score Y of a new user represented by $\mathbf{X}^\star$. The variables $\mathbf{X}_1^{(n)}, \ldots, \mathbf{X}_n^{(n)}$ model the database users' revealed preferences with respect to the first d items. They take values in $(\{0\} \cup [1, s])^d$, where a 0 at coordinate j of $\mathbf{X}_i^{(n)}$ means that the jth product has not been corated by both user i and the new user. The variable $\mathbf{X}^\star$ takes values in $(\{0\} \cup [1, s])^d$ and satisfies $\|\mathbf{X}^\star\| \geq 1$. The random variables $Y_1, \ldots, Y_n$ model users' ratings of the product of interest. They take

values in $[1, s]$ and, at time n, we only see a non-empty (random) subset of $\{Y_1, \ldots, Y_n\}$, indexed by $\mathcal{R}_n$.

The statistical problem with which we are faced is to estimate the regression function $\eta(\mathbf{x}^\star) = \mathbb{E}[Y|\mathbf{X}^\star = \mathbf{x}^\star]$. For this goal, we may use the database observations $(\mathbf{X}_1^{(n)}, Y_1), \ldots, (\mathbf{X}_n^{(n)}, Y_n)$ in order to construct an estimate $\eta_n(\mathbf{x}^\star)$ of $\eta(\mathbf{x}^\star)$. The approach we explore in this paper is a cosine-based k_n-nearest neighbor regression method, one of the most widely used algorithms in collaborative filtering (e.g., Sarwar et al. [15]).

Given $\mathbf{x}^\star \in (\{0\} \cup [1, s])^d - \mathbf{0}$ and the sample $(\mathbf{X}_1^{(n)}, Y_1), \ldots, (\mathbf{X}_n^{(n)}, Y_n)$, the idea of the cosine-type k_n-nearest neighbor (NN) regression method is to estimate $\eta(\mathbf{x}^\star)$ by a local averaging over those Y_i for which: (i) $\mathbf{X}_i^{(n)}$ is "close" to $\mathbf{x}^\star$ and (ii) $i \in \mathcal{R}_n$, that is, we effectively "see" the rating Y_i. For this, we scan through the k_n neighbors of $\mathbf{x}^\star$ among the database users $\mathbf{X}_i^{(n)}$ for which $i \in \mathcal{R}_n$ and estimate $\eta(\mathbf{x}^\star)$ by averaging the k_n corresponding Y_i. The closeness between users is assessed by a cosine-type similarity, defined for $\mathbf{x} = (x_1, \ldots, x_d)$ and $\mathbf{x}' = (x'_1, \ldots, x'_d)$ in $(\{0\} \cup [1, s])^d$ by

$$\bar{S}(\mathbf{x}, \mathbf{x}') = \frac{\sum_{j \in \mathcal{J}} x_j x'_j}{\sqrt{\sum_{j \in \mathcal{J}} x_j^2} \sqrt{\sum_{j \in \mathcal{J}} x_j'^2}},$$

where $\mathcal{J} = \{j \in \{1, \ldots, d\} : x_j \neq 0 \text{ and } x'_j \neq 0\}$ and, by convention, $\bar{S}(\mathbf{x}, \mathbf{x}') = 0$ if $\mathcal{J} = \emptyset$. To understand the rationale behind this proximity measure, just note that if $\mathcal{J} = \{1, \ldots, d\}$ then $\bar{S}(\mathbf{x}, \mathbf{x}')$ coincides with $\cos(\mathbf{x}, \mathbf{x}')$, i.e., two users are "close" with respect to $\bar{S}$ if their ratings are more or less proportional. However, the similarity $\bar{S}$, which will be used to measure the closeness between $\mathbf{X}^\star$ (the new user) and $\mathbf{X}_i^{(n)}$ (a database user) ignores possible non-answers in $\mathbf{X}^\star$ or $\mathbf{X}_i^{(n)}$, and is therefore more adapted to the recommendation setting. For example, in Table 1,

$$\bar{S}(\text{Bob, Jim}) = \bar{S}((0, 3, 3, 4, 5), (0, 6, 7, 8, 9)) = \bar{S}((3, 3, 4, 5), (6, 7, 8, 9)) \approx 0.99,$$

whereas

$$\bar{S}(\text{Bob, Lucy}) = \bar{S}((0, 3, 3, 4, 5), (3, 10, 2, 7, 0)) = \bar{S}((3, 3, 4), (10, 2, 7)) \approx 0.89.$$

Next, fix $\mathbf{x}^\star \in (\{0\} \cup [1, s])^d - \mathbf{0}$ and suppose to simplify that $M \subset M_i^{n+1-i}$ for each $i \in \mathcal{R}_n$. In this case, it is easy to see that $\mathbf{X}_i^{(n)} = \mathbf{X}_i^\star = (X_{i1}^\star, \ldots, X_{id}^\star)$, where

$$X_{ij}^\star = \begin{cases} X_{ij} & \text{if } j \in M \\ 0 & \text{otherwise.} \end{cases}$$

Besides, $Y_i \geq 1$,

$$\bar{S}(\mathbf{x}^\star, \mathbf{X}_i^\star) = \cos(\mathbf{x}^\star, \mathbf{X}_i^\star) > 0, \tag{2.2}$$

and an elementary calculation shows that the positive real number y which maximizes the similarity between $(\mathbf{x}^\star, y)$ and $(\mathbf{X}_i^\star, Y_i)$, that is

$$\bar{S}\left((\mathbf{x}^\star, y), (\mathbf{X}_i^\star, Y_i)\right) = \frac{\sum_{j\in M} x_j^\star X_{ij}^\star + yY_i}{\sqrt{\sum_{j\in M} x_j^{\star 2} + y^2}\sqrt{\sum_{j\in M} X_{ij}^{\star\, 2} + Y_i^2}},$$

is given by

$$y = \frac{\|\mathbf{x}^\star\|}{\|\mathbf{X}_i^\star\| \cos(\mathbf{x}^\star, \mathbf{X}_i^\star)}\, Y_i.$$

This suggests the following regression estimate $\eta_n(\mathbf{x}^\star)$ of $\eta(\mathbf{x}^\star)$:

$$\eta_n(\mathbf{x}^\star) = \|\mathbf{x}^\star\| \sum_{i\in\mathcal{R}_n} W_{ni}(\mathbf{x}^\star) \frac{Y_i}{\|\mathbf{X}_i^\star\|}, \tag{2.3}$$

where the integer k_n satisfies $1 \leq k_n \leq n$ and

$$W_{ni}(\mathbf{x}^\star) = \begin{cases} 1/k_n & \text{if } \mathbf{X}_i^\star \text{ is among the } k_n\text{-MS of } \mathbf{x}^\star \text{ in } \{\mathbf{X}_i^\star, i \in \mathcal{R}_n\} \\ 0 & \text{otherwise.} \end{cases}$$

In the above definition, the acronym "MS" (for Most Similar) means that we are searching for the k_n "closest" points of $\mathbf{x}^\star$ within the set $\{\mathbf{X}_i^\star, i \in \mathcal{R}_n\}$ using the similarity $\bar{S}$ — or, equivalently here, using the cosine proximity (by identity (2.2)). Note that the cosine term has been removed since it has asymptotically no influence on the estimate, as can be seen by a slight adaptation of the arguments of the proof of Lemma 6.1, Chapter 6, in Györfi et al. [9]. The estimate $\eta_n(\mathbf{x}^\star)$ is called the *cosine-type k_n-NN regression estimate* in the collaborative filtering literature. Now, recalling that definition (2.3) makes sense only when $M \subset M_i^{n+1-i}$ for each $i \in \mathcal{R}_n$ (that is, $\mathbf{X}_i^{(n)} = \mathbf{X}_i^\star$), the next step is to extend the definition of $\eta_n(\mathbf{x}^\star)$ to the general case. In view of (2.3), the most natural approach is to simply put

$$\eta_n(\mathbf{x}^\star) = \|\mathbf{x}^\star\| \sum_{i\in\mathcal{R}_n} W_{ni}(\mathbf{x}^\star) \frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}, \tag{2.4}$$

where

$$W_{ni}(\mathbf{x}^\star) = \begin{cases} 1/k_n & \text{if } \mathbf{X}_i^{(n)} \text{ is among the } k_n\text{-MS of } \mathbf{x}^\star \text{ in } \{\mathbf{X}_i^{(n)}, i \in \mathcal{R}_n\} \\ 0 & \text{otherwise.} \end{cases}$$

The acronym "MS" in the weight $W_{ni}(\mathbf{x}^\star)$ means that the k_n closest database points of $\mathbf{x}^\star$ are computed according to the similarity

$$S\left(\mathbf{x}^\star, \mathbf{X}_i^{(n)}\right) = p_i^{(n)} \bar{S}\left(\mathbf{x}^\star, \mathbf{X}_i^{(n)}\right), \quad \text{with } p_i^{(n)} = \frac{|M_i^{n+1-i} \cap M|}{|M|}$$

(here and throughout, notation $|A|$ means the cardinality of the finite set A). The factor $p_i^{(n)}$ in front of $\bar{S}$ is a penalty term which, roughly, avoids to over-promote the last users entering the database. Indeed, the effective number of items rated by these users will be eventually low and, consequently, their $\bar{S}$-proximity to $\mathbf{x}^\star$ will tend to remain high. On the other hand, for fixed i and n large enough, we know that $M \subset M_i^{n+1-i}$ and $\mathbf{X}_i^{(n)} = \mathbf{X}_i^\star$. This implies $p_i^{(n)} = 1$, $S(\mathbf{x}^\star, \mathbf{X}_i^{(n)}) = \bar{S}(\mathbf{x}^\star, \mathbf{X}_i^\star) = \cos(\mathbf{x}^\star, \mathbf{X}_i^\star)$ and shows that definition (2.4) generalizes definition (2.3). Therefore, we take the liberty to still call the estimate (2.4) the cosine-type k_n-NN regression estimate.

Remark 2.1 *A smoothed version of the similarity S could also be considered, typically*

$$S\left(\mathbf{x}^\star, \mathbf{X}_i^{(n)}\right) = \psi\left(p_i^{(n)}\right) \bar{S}\left(\mathbf{x}^\star, \mathbf{X}_i^{(n)}\right),$$

where $\psi : [0,1] \to [0,1]$ is a nondecreasing map satisfying $\psi(1/2) < 1$ (assuming $|M| \geq 2$). For example, the choice $\psi(p) = \sqrt{p}$ tends to promote users with a low number of rated items, provided the items corated by the new user are quite similar. In the present paper, we shall only consider the case $\psi(p) = p$, but the whole analysis carries over without difficulties for general functions ψ.

Remark 2.2 *Another popular approach to measure the closeness between users is the Pearson correlation coefficient. The extension of our results to Pearson-type similarities is not straightforward and more work is needed to address this challenging question. We refer the reader to Choi et al. [7] and Montaner et al. [12] for a comparative study and comments on the choice of the similarity.*

Finally, for definiteness of the estimate $\eta_n(\mathbf{x}^\star)$, some final remarks are in order:

(*i*) If $\mathbf{X}_i^{(n)}$ and $\mathbf{X}_j^{(n)}$ are equidistant from $\mathbf{x}^\star$, i.e., $S(\mathbf{x}^\star, \mathbf{X}_i^{(n)}) = S(\mathbf{x}^\star, \mathbf{X}_j^{(n)})$, then we have a tie and, for example, $\mathbf{X}_i^{(n)}$ may be declared "closer" to $\mathbf{x}^\star$ if $i < j$, that is, tie-breaking is done by indices.

(*ii*) If $|\mathcal{R}_n| < k_n$, then the weights $W_{ni}(\mathbf{x}^\star)$ are not defined. In this case, we conveniently set $W_{ni}(\mathbf{x}^\star) = 0$, i.e., $\eta_n(\mathbf{x}^\star) = 0$.

(*iii*) If $\mathbf{X}_i^{(n)} = \mathbf{0}$, then we take $W_{ni}(\mathbf{x}^\star) = 0$ and we adopt the convention $0 \times \infty = 0$ for the computation of $\eta_n(\mathbf{x}^\star)$.

(*iv*) With the above conventions, the identity $\sum_{i \in \mathcal{R}_n} W_{ni}(\mathbf{x}^\star) \leq 1$ holds in each case.

3 The regression function

Our objective in section 4 will be to establish consistency of the estimate $\eta_n(\mathbf{x}^\star)$ defined in (2.4) towards the regression function $\eta(\mathbf{x}^\star)$. To reach this goal, we first need to analyze the properties of $\eta(\mathbf{x}^\star)$. Surprisingly, the special form of $\eta_n(\mathbf{x}^\star)$ constrains the shape of $\eta(\mathbf{x}^\star)$. This is stated in Theorem 3.1 below.

Theorem 3.1 *Suppose that $\eta_n(\mathbf{X}^\star) \to \eta(\mathbf{X}^\star)$ in probability as $n \to \infty$. Then*

$$\eta(\mathbf{X}^\star) = \|\mathbf{X}^\star\| \, \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|} \Big| \frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} \right] \quad a.s.$$

Proof of Theorem 3.1. Recall that

$$\eta_n(\mathbf{X}^\star) = \|\mathbf{X}^\star\| \sum_{i \in \mathcal{R}_n} W_{ni}(\mathbf{X}^\star) \frac{Y_i}{\|\mathbf{X}_i^{(n)}\|},$$

and let

$$\varphi_n(\mathbf{X}^\star) = \sum_{i \in \mathcal{R}_n} W_{ni}(\mathbf{X}^\star) \frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}.$$

Since $(\eta_n(\mathbf{X}^\star))_n$ is a Cauchy sequence in probability and $\|\mathbf{X}^\star\| \geq 1$, $(\varphi_n(\mathbf{X}^\star))_n$ is also a Cauchy sequence. Thus, there exists a measurable function φ on $\mathbb{R}^d$ such that $\varphi_n(\mathbf{X}^\star) \to \varphi(\mathbf{X}^\star)$ in probability. Using the fact that $0 \leq \varphi_n(\mathbf{X}^\star) \leq s$ for all $n \geq 1$, we conclude that $0 \leq \varphi(\mathbf{X}^\star) \leq s$ a.s. as well.

Let us extract a sequence $(n_k)_k$ satisfying $\varphi_{n_k}(\mathbf{X}^\star) \to \varphi(\mathbf{X}^\star)$ a.s. Observing that, for $\mathbf{x}^\star \neq \mathbf{0}$,

$$\varphi_{n_k}(\mathbf{x}^\star) = \varphi_{n_k}\left(\frac{\mathbf{x}^\star}{\|\mathbf{x}^\star\|} \right),$$

we may write $\varphi(\mathbf{X}^\star) = \varphi(\mathbf{X}^\star / \|\mathbf{X}^\star\|)$ a.s. Consequently, the limit in probability of $(\eta_n(\mathbf{X}^\star))_n$ is

$$\|\mathbf{X}^\star\| \, \varphi\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} \right).$$

Therefore, by the uniqueness of the limit, $\eta(\mathbf{X}^\star) = \|\mathbf{X}^\star\|\,\varphi(\mathbf{X}^\star/\|\mathbf{X}^\star\|)$ a.s. Moreover,

$$\begin{aligned}
\varphi\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}\right) &= \mathbb{E}\left[\varphi\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}\right)\Big|\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}\right] \\
&= \mathbb{E}\left[\frac{\eta(\mathbf{X}^\star)}{\|\mathbf{X}^\star\|}\Big|\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|}\Big|\mathbf{X}^\star\right]\Big|\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}\right] \\
&= \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|}\Big|\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}\right],
\end{aligned}$$

since $\sigma(\mathbf{X}^\star/\|\mathbf{X}^\star\|) \subset \sigma(\mathbf{X}^\star)$. This concludes the proof of the theorem.

□

An important consequence of Theorem 3.1 is that if we intend to prove any consistency result regarding the estimate $\eta_n(\mathbf{x}^\star)$, then we have to assume that the regression function $\eta(\mathbf{x}^\star)$ has the special form

$$\eta(\mathbf{x}^\star) = \|\mathbf{x}^\star\|\varphi(\mathbf{x}^\star), \quad \text{where} \quad \varphi(\mathbf{x}^\star) = \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|}\Big|\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} = \frac{\mathbf{x}^\star}{\|\mathbf{x}^\star\|}\right] \quad (\mathbf{F}).$$

This will be our fundamental requirement throughout the paper, and it will be denoted by (**F**). In particular, if $\tilde{\mathbf{x}}^\star = \lambda\mathbf{x}^\star$ with $\lambda > 0$, then $\eta(\tilde{\mathbf{x}}^\star) = \lambda\eta(\mathbf{x}^\star)$. That is, if two ratings $\mathbf{x}^\star$ and $\tilde{\mathbf{x}}^\star$ are proportional, then so must be the values of the regression function at $\mathbf{x}^\star$ and $\tilde{\mathbf{x}}^\star$, respectively.

4 Consistency

In this section, we establish the L_1 consistency of the regression estimate $\eta_n(\mathbf{x}^\star)$ towards the regression function $\eta(\mathbf{x}^\star)$. Using L_1 consistency is essentially a matter of taste, and all the subsequent results may be easily adapted to L_p norms without too much effort. In the proofs, we will make repeated use of the two following facts. Recall that, for a fixed $i \in \mathcal{R}_n$, the random variable $\mathbf{X}_i^\star = (X_{i1}^\star, \ldots, X_{id}^\star)$ is defined by

$$X_{ij}^\star = \begin{cases} X_{ij} & \text{if } j \in M \\ 0 & \text{otherwise,} \end{cases}$$

and $\mathbf{X}_i^{(n)} = \mathbf{X}_i^\star$ as soon as $M \subset M_i^{n+1-i}$. Recall also that, by definition, $\|\mathbf{X}_i^\star\| \geq 1$.

Fact 4.1 *For each $i \in \mathcal{R}_n$,*

$$S(\mathbf{X}^\star, \mathbf{X}_i^\star) = \bar{S}(\mathbf{X}^\star, \mathbf{X}_i^\star) = \cos(\mathbf{X}^\star, \mathbf{X}_i^\star) = 1 - \frac{1}{2}\,\mathrm{d}^2\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}, \frac{\mathbf{X}_i^\star}{\|\mathbf{X}_i^\star\|}\right),$$

where d *is the usual Euclidean distance on $\mathbb{R}^d$.*

Fact 4.2 *Let, for all $i \geq 1$,*

$$T_i = \min(k \geq i : M_i^{k+1-i} \supset M)$$

be the first time instant when user i has rated all the films indexed by M. Set

$$\mathcal{L}_n = \{i \in \mathcal{R}_n : T_i \leq n\}, \tag{4.1}$$

and define, for $i \in \mathcal{L}_n$,

$$W_{ni}^\star(\mathbf{x}^\star) = \begin{cases} 1/k_n & \text{if } \mathbf{X}_i^\star \text{ is among the } k_n\text{-MS of } \mathbf{x}^\star \text{ in } \{\mathbf{X}_i^\star, i \in \mathcal{L}_n\} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$W_{ni}^\star(\mathbf{x}^\star) = \begin{cases} 1/k_n & \text{if } \frac{\mathbf{X}_i^\star}{\|\mathbf{X}_i^\star\|} \text{ is among the } k_n\text{-NN of } \frac{\mathbf{x}^\star}{\|\mathbf{x}^\star\|} \text{ in } \left\{\frac{\mathbf{X}_i^\star}{\|\mathbf{X}_i^\star\|}, i \in \mathcal{L}_n\right\} \\ 0 & \text{otherwise,} \end{cases}$$

where the k_n-NN are evaluated with respect to the Euclidean distance on $\mathbb{R}^d$. That is, the $W_{ni}^\star(\mathbf{x}^\star)$ are the usual Euclidean NN weights (Györfi et al. [9]), indexed by the random set $\mathcal{L}_n$.

Recall that $|\mathcal{R}_n|$ represents the number of users who have already provided information about the variable of interest (the movie Titanic in our example) at time n. We are now in a position to state the main result of this section.

Theorem 4.1 *Suppose that $|M| \geq 2$ and that assumption* (**F**) *is satisfied. Suppose that $k_n \to \infty$, $|\mathcal{R}_n| \to \infty$ a.s. and $\mathbb{E}[k_n/|\mathcal{R}_n|] \to 0$ as $n \to \infty$. Then*

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| \to 0 \quad \text{as } n \to \infty.$$

Thus, to achieve consistency, the number of nearest neighbors k_n, over which one averages in order to estimate the regression function, should on the one hand tend to infinity but should, on the other hand, be small with respect to the cardinality of the subset of database users who have already rated the item of interest. We illustrate this result by working out two examples.

Example 4.1 *Consider, to start with, the somewhat ideal situation where all users in the database have rated the item of interest. In this case,* $\mathcal{R}_n = \{1, \ldots, n\}$*, and the asymptotic conditions on* k_n *become* $k_n \to \infty$ *and* $k_n/n \to 0$ *as* $n \to \infty$*. These are just the well-known conditions ensuring consistency of the usual (i.e., Euclidean) NN regression estimate (Györfi et al. [9], Chapter 6).*

Example 4.2 *In this more sophisticated model, we recursively define the sequence* $(\mathcal{R}_n)_n$ *as follows. Fix, for simplicity,* $\mathcal{R}_1 = \{1\}$*. At step* $n \geq 2$*, we first decide or not to add one element to* $\mathcal{R}_{n-1}$ *with probability* $p \in (0,1)$*, independently of the data. If we decide to increase* $\mathcal{R}_n$*, then we do it by picking a random variable* B_n *uniformly over the set* $\{1, \ldots, n\} - \mathcal{R}_{n-1}$*, and set* $\mathcal{R}_n = \mathcal{R}_{n-1} \cup \{B_n\}$*; otherwise,* $\mathcal{R}_n = \mathcal{R}_{n-1}$*. Clearly,* $|\mathcal{R}_n| - 1$ *is a sum of* $n-1$ *independent Bernoulli random variables with parameter* p*, and it has therefore a binomial distribution with parameters* $n-1$ *and* p*. Consequently,*

$$\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] = \frac{k_n\left[1-(1-p)^n\right]}{np}.$$

In this setting, consistency holds provided $k_n \to \infty$ *and* $k_n = o(n)$ *as* $n \to \infty$*.*

In the sequel, the letter C will denote a positive constant, the value of which may vary from line to line. Proof of Theorem 4.1 will strongly rely on facts 4.1, 4.2 and the following proposition.

Proposition 4.1 *Suppose that* $|M| \geq 2$ *and that assumption* $(\mathbf{F})$ *is satisfied. Let* $\alpha_{ni} = \mathbb{P}(M^{n+1-i} \not\supset M \mid M)$*. Then*

$$\begin{aligned}
&\mathbb{E}\left|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)\right| \\
&\quad \leq C\Bigg\{\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] + \mathbb{E}\left[\frac{1}{|\mathcal{R}_n|}\sum_{i\in\mathcal{R}_n}\mathbb{E}\alpha_{ni}\right] + \mathbb{E}\left[\prod_{i\in\mathcal{R}_n}\alpha_{ni}\right] \\
&\qquad + \mathbb{E}\Big|\sum_{i\in\mathcal{L}_n} W_{ni}^\star(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^\star\|} - \varphi(\mathbf{X}^\star)\Big|\Bigg\},
\end{aligned}$$

where $\mathcal{R}_n$ *stands for the non-empty subset of users who have already provided information about the variable of interest at time* n *and* $\mathcal{L}_n$ *is defined in* (4.1)*.*

Proof of Proposition 4.1. Since $\|\mathbf{X}^\star\| \leq s\sqrt{d}$, it will be enough to upper bound the quantity

$$\mathbb{E}\left|\sum_{i\in\mathcal{R}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|} - \varphi(\mathbf{X}^\star)\right|.$$

To this aim, we write

$$\mathbb{E}\left|\sum_{i\in\mathcal{R}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|} - \varphi(\mathbf{X}^\star)\right|$$
$$\leq \mathbb{E}\left[\sum_{i\in\mathcal{L}_n^c} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}\right] + \mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|} - \varphi(\mathbf{X}^\star)\right|,$$

where the symbol A^c denotes the complement of the set A. Let the event

$$\mathcal{A}_n = \left[\exists i \in \mathcal{L}_n^c : \mathbf{X}_i^{(n)} \text{ is among the } k_n\text{-MS of } \mathbf{X}^\star \text{ in } \{\mathbf{X}_i^{(n)}, i \in \mathcal{R}_n\}\right].$$

Since $\sum_{i\in\mathcal{L}_n^c} W_{ni}(\mathbf{X}^\star) \leq 1$, we have

$$\mathbb{E}\left[\sum_{i\in\mathcal{L}_n^c} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}\right] = \mathbb{E}\left[\sum_{i\in\mathcal{L}_n^c} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|}\mathbf{1}_{\mathcal{A}_n}\right] \leq s\mathbb{P}(\mathcal{A}_n).$$

Observing that, for $i \in \mathcal{L}_n$, $\mathbf{X}_i^{(n)} = \mathbf{X}_i^\star$ and $W_{ni}(\mathbf{X}^\star)\mathbf{1}_{\mathcal{A}_n^c} = W_{ni}^\star(\mathbf{X}^\star)\mathbf{1}_{\mathcal{A}_n^c}$ (fact 4.2), we obtain

$$\begin{aligned}
&\mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^{(n)}\|} - \varphi(\mathbf{X}^\star)\right| \\
&= \mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^\star\|} - \varphi(\mathbf{X}^\star)\right| \\
&= \mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^\star\|} - \varphi(\mathbf{X}^\star)\right|\mathbf{1}_{\mathcal{A}_n} + \mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}^\star(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^\star\|} - \varphi(\mathbf{X}^\star)\right|\mathbf{1}_{\mathcal{A}_n^c} \\
&\leq s\mathbb{P}(\mathcal{A}_n) + \mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}^\star(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^\star\|} - \varphi(\mathbf{X}^\star)\right|.
\end{aligned}$$

Applying finally Lemma 6.5 completes the proof of the proposition.

□

We are now in a position to prove Theorem 4.1.

Proof of Theorem 4.1. According to Proposition 4.1, Lemma 6.1 and Lemma 6.2, the result will be proven if we show that

$$\mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}^\star(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^\star\|} - \varphi(\mathbf{X}^\star)\right| \to 0 \quad \text{as} \quad n \to \infty.$$

For $L_n \in \mathcal{P}(\{1,\dots,n\})$, set

$$Z^n_{L_n} = \frac{1}{k_n}\sum_{i\in L_n} \mathbf{1}_{\left[\frac{\mathbf{X}^\star_i}{\|\mathbf{X}^\star_i\|} \text{ is among the } k_n\text{-NN of } \frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} \text{ in } \left\{\frac{\mathbf{X}^\star_i}{\|\mathbf{X}^\star_i\|}, i\in L_n\right\}\right]} \frac{Y_i}{\|\mathbf{X}^\star_i\|} - \varphi(\mathbf{X}^\star).$$

Conditionally on the event $[M=m]$, the random variables $\mathbf{X}^\star$ and $\{\mathbf{X}^\star_i, i \in L_n\}$ are independent and identically distributed. Thus, applying Theorem 6.1 in [9], we obtain

$$\forall \varepsilon > 0,\ \exists A_m \geq 1 : k_n \geq A_m \text{ and } \frac{|L_n|}{k_n} \geq A_m \Longrightarrow \mathbb{E}_m|Z^n_{L_n}| \leq \varepsilon,$$

where we use the notation $\mathbb{E}_m[.] = \mathbb{E}[.|M=m]$. Let $\mathbb{P}_m(.) = \mathbb{P}(.|M=m)$. By independence,

$$\mathbb{E}_m|Z^n_{\mathcal{L}_n}| = \sum_{L_n\in\mathcal{P}(\{1,\dots,n\})} \mathbb{E}_m|Z^n_{L_n}|\,\mathbb{P}_m(\mathcal{L}_n = L_n).$$

Consequently, letting $A = \max A_m$, where the maximum is taken over all possible choices of $m \in \mathcal{P}^\star(\{1,\dots,d\})$ we get, for all n such that $k_n \geq A$,

$$\begin{aligned}\mathbb{E}_m|Z^n_{\mathcal{L}_n}| &= \sum_{\substack{L_n\in\mathcal{P}(\{1,\dots,n\})\\ |L_n|\geq Ak_n}} \mathbb{E}_m|Z^n_{L_n}|\,\mathbb{P}_m(\mathcal{L}_n = L_n)\\ &\quad + \sum_{\substack{L_n\in\mathcal{P}(\{1,\dots,n\})\\ |L_n|< Ak_n}} \mathbb{E}_m|Z^n_{L_n}|\,\mathbb{P}_m(\mathcal{L}_n = L_n)\\ &\leq \varepsilon + s\mathbb{P}_m(|\mathcal{L}_n| < Ak_n).\end{aligned}$$

Therefore

$$\mathbb{E}|Z^n_{\mathcal{L}_n}| = \mathbb{E}\left[\mathbb{E}\left[|Z^n_{\mathcal{L}_n}| \,\middle|\, M\right]\right] \leq \varepsilon + s\mathbb{P}\left(|\mathcal{L}_n| < Ak_n\right).$$

Moreover, by Lemma 6.2,

$$\frac{|\mathcal{L}_n|}{k_n} = \frac{|\mathcal{R}_n|}{k_n}\left(1 - \frac{|\mathcal{L}^c_n|}{|\mathcal{R}_n|}\right) \to \infty \text{ in probability as } n\to\infty.$$

Thus, for all $\varepsilon > 0$, $\limsup_{n\to\infty}\mathbb{E}|Z^n_{\mathcal{L}_n}| \leq \varepsilon$, whence $\mathbb{E}|Z^n_{\mathcal{L}_n}| \to 0$ as $n\to\infty$. This shows the desired result.

□

5 Rates of convergence

In this section, we bound the rate of convergence of $\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)|$ for the cosine-type k_n-NN regression estimate. To reach this objective, we will require that the function

$$\varphi(\mathbf{x}^\star) = \mathbb{E}\left[\frac{Y}{\|\mathbf{X}^\star\|}\Big|\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|} = \frac{\mathbf{x}^\star}{\|\mathbf{x}^\star\|}\right]$$

satisfies a Lipschitz-type property with respect to the similarity $\bar{S}$. More precisely, we say that φ is Lipschitz with respect to $\bar{S}$ if there exists a constant $C > 0$ such that, for all $\mathbf{x}$ and $\mathbf{x}'$ in $\mathbb{R}^d$,

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{x}')| \leq C\sqrt{1 - \bar{S}(\mathbf{x}, \mathbf{x}')}.$$

In particular, for $\mathbf{x}$ and $\mathbf{x}' \in \mathbb{R}^d - \mathbf{0}$ with the same null components, this property can be rewritten as

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{x}')| \leq \frac{C}{\sqrt{2}}\,\mathrm{d}\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}, \frac{\mathbf{x}'}{\|\mathbf{x}'\|}\right),$$

where we recall that d denotes Euclidean distance.

Theorem 5.1 *Suppose that assumption* (**F**) *is satisfied and that φ is Lipschitz with respect to $\bar{S}$. Let $\alpha_{ni} = \mathbb{P}(M^{n+1-i} \not\supset M \,|\, M)$, and assume that $|M| \geq 4$. Then there exists $C > 0$ such that, for all $n \geq 1$,*

$$\begin{aligned}&\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| \\ &\quad\leq C\left\{\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\sum_{i\in\mathcal{R}_n}\mathbb{E}\,\alpha_{ni}\right] + \mathbb{E}\left[\prod_{i\in\mathcal{R}_n}\alpha_{ni}\right] + \mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{P_n}\right] + \frac{1}{\sqrt{k_n}}\right\},\end{aligned}$$

where $P_n = 1/(|M| - 1)$ if $k_n \leq |\mathcal{R}_n|$, and $P_n = 1$ otherwise.

To get an intuition on the meaning of Theorem 5.1, it helps to note that the terms depending on α_{ni} do measure the influence of the unrated items on the performance of the estimate. Clearly, this performance improves as the α_{ni} decrease, i.e., as the proportion of rated items growths. On the other hand, the term $\mathbb{E}[(k_n/|\mathcal{R}_n|)^{P_n}]$ can be interpreted as a bias term in dimension $|M| - 1$, whereas $1/\sqrt{k_n}$ represents a variance term. As usual in nonparametric estimation, the rate of convergence of the estimate is dramatically deteriorated as $|M|$ becomes large. However, in practice, this drawback may be circumvented by using preliminary dimension reduction steps, such as factorial methods (PCA, etc.) or inverse regression methods (SIR, etc.).

Example 5.1 *(**cont. Example 4.1**) Recall that we assume, in this ideal model, that $\mathcal{R}_n = \{1, \ldots, n\}$. Suppose in addition that $M = \{1, \ldots, d\}$, i.e., any new user in the database rates all products the first time he enters the database. Then the upper bound of Theorem 5.1 becomes*

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| = \mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/(d-1)} + \frac{1}{\sqrt{k_n}}\right).$$

Since neither $\mathcal{R}_n$ nor M are random in this model, we see that there is no influence of the dynamical rating process. Besides, we recognize the usual rate of convergence of the Euclidean NN regression estimate (Györfi et al. [9], Chapter 6) in dimension $d-1$. In particular, the choice $k_n \sim n^{2/(d+1)}$ leads to

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| = \mathrm{O}\left(n^{-1/(d+1)}\right).$$

Note that we are led to a $d-1$-dimensional rate of convergence (instead of the usual d) just because everything happens as if the data is projected on the unit sphere of $\mathbb{R}^d$.

Example 5.2 *(**cont. Example 4.2**) In addition to model 4.2, we suppose that at each time, a user entering the game reveals his preferences according to the following sequential procedure. At time 1, the user rates exactly 4 items by randomly guessing in $\{1, \ldots, d\}$. At time 2, he updates his preferences by adding exactly one rating among his unrated items, randomly chosen in $\{1, \ldots, d\} - M_1^1$. Similarly, at time 3, the user revises his preferences according to a new item uniformly selected in $\{1, \ldots, d\} - M_1^2$, and so on. In such a scenario, $|M^j| = \min(d, j+3)$ and thus, $M^j = \{1, \ldots, d\}$ for $j \geq d-3$. Moreover, since $|M| = 4$, a moment's thought shows that*

$$\alpha_{ni} = \begin{cases} 0 & \text{if } i \leq n-d+4 \\ \displaystyle 1 - \frac{\binom{d-4}{n-i}}{\binom{d}{n+4-i}} & \text{if } n-d+5 \leq i \leq n. \end{cases}$$

Assuming $n \geq d-5$, we obtain

$$\begin{aligned} \sum_{i \in \mathcal{R}_n} \alpha_{ni} &\leq \sum_{i=n-d+5}^{n} \alpha_{ni} \\ &\leq \sum_{i=n-d+5}^{n} \left(1 - \frac{(n+4-i)(n+3-i)(n+2-i)(n+1-i)}{d(d-1)(d-2)(d-3)}\right) \\ &\leq (d-4)\left(1 - \frac{24}{d(d-1)(d-2)(d-3)}\right). \end{aligned}$$

Similarly, letting $\mathcal{R}_{n0} = \mathcal{R}_n \cap \{n-d+5, \ldots, n\}$, *we have*

$$\begin{aligned}\prod_{i\in\mathcal{R}_n} \alpha_{ni} &= \prod_{i\in\mathcal{R}_{n0}} \alpha_{ni}\, \mathbf{1}_{\{\min(\mathcal{R}_n)\geq n-d+5\}} \\ &\leq \left(1 - \frac{24}{d(d-1)(d-2)(d-3)}\right)^{|\mathcal{R}_{n0}|} \mathbf{1}_{\{\min(\mathcal{R}_n)\geq n-d+5\}}.\end{aligned}$$

Since $|\mathcal{R}_n| - 1$ *has binomial distribution with parameters* $n-1$ *and* p, *we obtain*

$$\begin{aligned}\mathbb{E}\left[\prod_{i\in\mathcal{R}_n} \alpha_{ni}\right] &\leq \mathbb{P}\big(\min(\mathcal{R}_n) \geq n-d+5\big) \\ &\leq \mathbb{P}\big(|\mathcal{R}_n| \leq d-5\big) \leq \frac{C}{n}.\end{aligned}$$

Finally, applying Jensen's inequality,

$$\begin{aligned}\mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{P_n}\right] &= \mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{1/3} \mathbf{1}_{\{k_n\leq|\mathcal{R}_n|\}}\right] + \mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\mathbf{1}_{\{k_n>|\mathcal{R}_n|\}}\right] \\ &\leq C\left(\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right]\right)^{1/3} \leq C\left(\frac{k_n}{n}\right)^{1/3}.\end{aligned}$$

Putting all the pieces together, we get with Theorem 5.1

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| = \mathrm{O}\left(\left(\frac{k_n}{n}\right)^{1/3} + \frac{1}{\sqrt{k_n}}\right).$$

In particular, the choice $k_n \sim n^{2/5}$ *leads to*

$$\mathbb{E}\,|\eta_n(\mathbf{X}^\star) - \eta(\mathbf{X}^\star)| = \mathrm{O}(n^{-1/5}),$$

which is the usual NN regression estimate rate of convergence when the data is projected on the unit sphere of $\mathbb{R}^4$.

Proof of Theorem 5.1. Starting from Proposition 4.1, we just need to upper bound the quantity

$$\mathbb{E}\left|\sum_{i\in\mathcal{L}_n} W_{ni}^\star(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}_i^\star\|} - \varphi(\mathbf{X}^\star)\right|.$$

A combination of Lemma 6.6 and the proof of Theorem 6.2 in [9] shows that

$$\mathbb{E}\left| \sum_{i\in\mathcal{L}_n} W^\star_{ni}(\mathbf{X}^\star)\frac{Y_i}{\|\mathbf{X}^\star_i\|} - \varphi(\mathbf{X}^\star)\right| \\ \leq C\left\{\frac{1}{\sqrt{k_n}} + \mathbb{E}\left[\left(\frac{k_n}{|\mathcal{L}_n|}\right)^{1/(|M|-1)}\mathbf{1}_{\{\mathcal{L}_n\neq\emptyset\}}\right] + \mathbb{P}(\mathcal{L}_n=\emptyset)\right\}. \tag{5.1}$$

We obtain

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{k_n}{|\mathcal{L}_n|}\right)^{1/(|M|-1)}\mathbf{1}_{\{\mathcal{L}_n\neq\emptyset\}}\right] \\
&\quad= \mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|\big(1-|\mathcal{L}_n^c|/|\mathcal{R}_n|\big)}\right)^{1/(|M|-1)}\mathbf{1}_{\{|\mathcal{L}_n^c|\leq|\mathcal{R}_n|/2\}}\right] \\
&\qquad+ \mathbb{E}\left[\left(\frac{k_n}{|\mathcal{L}_n|}\right)^{1/(|M|-1)}\mathbf{1}_{\{|\mathcal{L}_n^c|>|\mathcal{R}_n|/2\}}\mathbf{1}_{\{\mathcal{L}_n\neq\emptyset\}}\right] \\
&\quad\leq \mathbb{E}\left[\left(\frac{2k_n}{|\mathcal{R}_n|}\right)^{1/(|M|-1)}\right] + \mathbb{E}\left[k_n^{1/(|M|-1)}\mathbf{1}_{\{|\mathcal{L}_n^c|>|\mathcal{R}_n|/2\}}\right].
\end{aligned}$$

Since $|M|\geq 4$, one has $2^{1/(|M|-1)}\leq 2$ and $k_n^{1/(|M|-1)}\leq k_n$ in the rightmost term, so that, thanks to Lemma 6.2,

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{k_n}{|\mathcal{L}_n|}\right)^{1/(|M|-1)}\mathbf{1}_{\{\mathcal{L}_n\neq\emptyset\}}\right] \\
&\quad\leq C\left\{\mathbb{E}\left[\left(\frac{k_n}{|\mathcal{R}_n|}\right)^{1/(|M|-1)}\right] + \mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\sum_{i\in\mathcal{R}_n}\mathbb{E}\,\alpha_{ni}\right]\right\}.
\end{aligned}$$

The theorem is a straightforward combination of Proposition 4.1, inequality (5.1), and Lemma 6.1.

□

6 Technical lemmas

Before stating some technical lemmas, we remind the reader that $\mathcal{R}_n$ stands for the non-empty subset of $\{1,\ldots,n\}$ of users who have already rated the variable of interest at time n. Recall also that, for all $i\geq 1$,

$$T_i = \min(k\geq i : M_i^{k+1-i}\supset M)$$

and

$$\mathcal{L}_n = \{i \in \mathcal{R}_n : T_i \leq n\}.$$

Lemma 6.1 *We have*

$$\mathbb{P}(\mathcal{L}_n = \emptyset) = \mathbb{E}\left[\prod_{i \in \mathcal{R}_n} \alpha_{ni}\right] \to 0 \quad \textit{as } n \to \infty.$$

Proof of Lemma 6.1. Conditionally on M and $\mathcal{R}_n$, the random variables $\{T_i, i \in \mathcal{R}_n\}$ are independent. Moreover, the sequence $(M^n)_{n \geq 1}$ is nondecreasing. Thus, the identity $[T_i > n] = [M_i^{n+1-i} \not\supset M]$ holds for all $i \in \mathcal{R}_n$. Hence,

$$\begin{aligned}
\mathbb{P}(\mathcal{L}_n = \emptyset) &= \mathbb{P}\left(\forall i \in \mathcal{R}_n : T_i > n\right) \\
&= \mathbb{E}\left[\mathbb{P}\left(\forall i \in \mathcal{R}_n : T_i > n \,\middle|\, \mathcal{R}_n, M\right)\right] \\
&= \mathbb{E}\left[\prod_{i \in \mathcal{R}_n} \mathbb{P}\left(T_i > n \,\middle|\, \mathcal{R}_n, M\right)\right] \\
&= \mathbb{E}\left[\prod_{i \in \mathcal{R}_n} \mathbb{P}\left(M_i^{n+1-i} \not\supset M \,\middle|\, M\right)\right] \\
&\quad \text{(by independence of } (M_i^{n+1-i}, M) \text{ and } \mathcal{R}_n \\
&= \mathbb{E}\left[\prod_{i \in \mathcal{R}_n} \alpha_{ni}\right].
\end{aligned}$$

The last statement of the lemma is clear since, for all i, $\alpha_{ni} \to 0$ a.s. as $n \to \infty$.

□

Lemma 6.2 *We have*

$$\mathbb{E}\left[\frac{|\mathcal{L}_n^c|}{|\mathcal{R}_n|}\right] = \mathbb{E}\left[\frac{1}{|\mathcal{R}_n|} \sum_{i \in \mathcal{R}_n} \mathbb{E}\, \alpha_{ni}\right]$$

and

$$\mathbb{E}\left[\frac{1}{|\mathcal{L}_n|} \mathbf{1}_{\{\mathcal{L}_n \neq \emptyset\}}\right] \leq 2\mathbb{E}\left[\frac{1}{|\mathcal{R}_n|}\right] + 2\mathbb{E}\left[\frac{1}{|\mathcal{R}_n|} \sum_{i \in \mathcal{R}_n} \mathbb{E}\, \alpha_{ni}\right].$$

Moreover, if $\lim_{n \to \infty} |\mathcal{R}_n| = \infty$ *a.s., then*

$$\lim_{n \to \infty} \mathbb{E}\left[\frac{|\mathcal{L}_n^c|}{|\mathcal{R}_n|}\right] = 0.$$

Proof of Lemma 6.2. First, using the fact that the sequence $(M^n)_{n\geq 1}$ is nondecreasing, we see that for all $i \in \mathcal{R}_n$, $[T_i > n] = [M_i^{n+1-i} \not\supset M]$. Next, recalling that $\mathcal{R}_n$ is independent of T_i for fixed i, we obtain

$$\mathbb{E}\left[\frac{|\mathcal{L}_n^c|}{|\mathcal{R}_n|}\,\Big|\,\mathcal{R}_n\right] = \frac{1}{|\mathcal{R}_n|}\mathbb{E}\left[\sum_{i\in\mathcal{R}_n}\mathbf{1}_{\{T_i>n\}}\,\Big|\,\mathcal{R}_n\right] = \frac{1}{|\mathcal{R}_n|}\sum_{i\in\mathcal{R}_n}\mathbb{P}(M_i^{n+1-i}\not\supset M),$$

and this proves the first statement of the lemma. Now define $\mathcal{J}_n = \{n+1-i, i\in\mathcal{R}_n\}$ and observe that

$$\mathbb{E}\left[\frac{|\mathcal{L}_n^c|}{|\mathcal{R}_n|}\right] = \mathbb{E}\left[\frac{1}{|\mathcal{J}_n|}\sum_{j\in\mathcal{J}_n}\mathbb{P}(M^j\not\supset M)\right],$$

where we used $|\mathcal{J}_n| = |\mathcal{R}_n|$. Since, by assumption, $|\mathcal{J}_n| = |\mathcal{R}_n| \to \infty$ a.s. as $n\to\infty$ and $\mathbb{P}(M^j\not\supset M)\to 0$ as $j\to\infty$, we obtain

$$\lim_{n\to\infty}\frac{1}{|\mathcal{J}_n|}\sum_{j\in\mathcal{J}_n}\mathbb{P}(M^j\not\supset M) = 0 \quad \text{a.s.}$$

The conclusion follows by applying Lebesgue's dominated convergence Theorem. The second statement of the lemma is obtained from the following chain of inequalities:

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{|\mathcal{L}_n|}\mathbf{1}_{\{\mathcal{L}_n\neq\emptyset\}}\right] &= \mathbb{E}\left[\frac{1}{|\mathcal{R}_n|\big(1-|\mathcal{L}_n^c|/|\mathcal{R}_n|\big)}\mathbf{1}_{\{\mathcal{L}_n\neq\emptyset\}}\right]\\
&= \mathbb{E}\left[\frac{1}{|\mathcal{R}_n|\big(1-|\mathcal{L}_n^c|/|\mathcal{R}_n|\big)}\mathbf{1}_{\{|\mathcal{L}_n^c|\leq|\mathcal{R}_n|/2\}}\right]\\
&\quad +\mathbb{E}\left[\frac{1}{|\mathcal{L}_n|}\mathbf{1}_{\{|\mathcal{L}_n^c|>|\mathcal{R}_n|/2\}}\mathbf{1}_{\{\mathcal{L}_n\neq\emptyset\}}\right]\\
&\leq 2\mathbb{E}\left[\frac{1}{|\mathcal{R}_n|}\right] + \mathbb{P}\left(|\mathcal{L}_n^c| > \frac{|\mathcal{R}_n|}{2}\right)\\
&\leq 2\mathbb{E}\left[\frac{1}{|\mathcal{R}_n|}\right] + 2\mathbb{E}\left[\frac{|\mathcal{L}_n^c|}{|\mathcal{R}_n|}\right].
\end{aligned}$$

Applying the first part of the lemma completes the proof.

□

Lemma 6.3 *Denote by* $\mathbf{Z}^\star$ *and* $\mathbf{Z}_1^\star$ *the random variables* $\mathbf{Z}^\star = \mathbf{X}^\star/\|\mathbf{X}^\star\|$, $\mathbf{Z}_1^\star = \mathbf{X}_1^\star/\|\mathbf{X}_1^\star\|$, *and let* $\xi(\mathbf{Z}^\star) = \mathbb{P}(S(\mathbf{Z}^\star, \mathbf{Z}_1^\star) > 1/2 \,|\, \mathbf{Z}^\star)$. *Then*

$$\mathbb{P}\Big(2k_n > |\mathcal{L}_n|\xi(\mathbf{Z}^\star) \,\Big|\, \mathcal{L}_n, M\Big) \leq 2\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|} \,\middle|\, \mathcal{L}_n\right] \mathbb{E}\left[\frac{1}{\xi(\mathbf{Z}^\star)} \,\middle|\, M\right] + \mathbb{E}\left[\frac{|\mathcal{L}_n^c|}{|\mathcal{R}_n|} \,\middle|\, \mathcal{L}_n, M\right].$$

Proof of Lemma 6.3. If M is fixed, $\mathbf{Z}^\star$ is independent of $\mathcal{L}_n$ and $\mathcal{R}_n$. Thus, by Markov's inequality,

$$\begin{aligned}
&\mathbb{P}\Big(2k_n > |\mathcal{L}_n|\xi(\mathbf{Z}^\star) \,\Big|\, \mathcal{L}_n, M, \mathcal{R}_n\Big) \\
&\quad = \mathbb{P}\Big(2k_n > |\mathcal{R}_n|\xi(\mathbf{Z}^\star) - |\mathcal{L}_n^c|\xi(\mathbf{Z}^\star) \,\Big|\, \mathcal{L}_n, M, \mathcal{R}_n\Big) \\
&\quad = \mathbb{P}\Big(2k_n + |\mathcal{L}_n^c|\xi(\mathbf{Z}^\star) \geq |\mathcal{R}_n|\xi(\mathbf{Z}^\star) \,\Big|\, \mathcal{L}_n, M, \mathcal{R}_n\Big) \\
&\quad \leq \frac{2k_n}{|\mathcal{R}_n|}\mathbb{E}\left[\frac{1}{\xi(\mathbf{Z}^\star)} \,\middle|\, M\right] + \frac{|\mathcal{L}_n^c|}{|\mathcal{R}_n|}.
\end{aligned}$$

The proof is completed by observing that $\mathcal{R}_n$ and M are independent random variables.

□

Let $\mathcal{B}(\mathbf{x}, \varepsilon)$ be the closed Euclidean ball in $\mathbb{R}^d$ centered at $\mathbf{x}$ of radius ε. Recall that the support of a probability measure μ is defined as the closure of the collection of all $\mathbf{x}$ with $\mu(\mathcal{B}(\mathbf{x}, \varepsilon)) > 0$ for all $\varepsilon > 0$. The next lemma can be proved with a slight modification of the proof of Lemma 10.2 in Devroye et al. [8].

Lemma 6.4 *Let* μ *be a probability measure on* $\mathbb{R}^d$ *with a compact support. Then*

$$\int \frac{1}{\mu(\mathcal{B}(\mathbf{x}, r))}\mu(\mathrm{d}\mathbf{x}) \leq C,$$

with $C > 0$ *a constant depending upon* d *and* r *only.*

Lemma 6.5 *Suppose that* $|M| \geq 2$, *and let the event*

$$\mathcal{A}_n = \Big[\exists i \in \mathcal{L}_n^c : \mathbf{X}_i^{(n)} \text{ is among the } k_n\text{-MS of } \mathbf{X}^\star \text{ in } \{\mathbf{X}_i^{(n)}, i \in \mathcal{R}_n\}\Big].$$

Then

$$\mathbb{P}(\mathcal{A}_n) \leq C\left\{\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] + \mathbb{E}\left[\frac{1}{|\mathcal{R}_n|}\sum_{i\in\mathcal{R}_n}\mathbb{E}\alpha_{ni}\right] + \mathbb{E}\left[\prod_{i\in\mathcal{R}_n}\alpha_{ni}\right]\right\}.$$

Proof of Lemma 6.5. Recall that, for a fixed $i \in \mathcal{R}_n$, the random variable $\mathbf{X}_i^\star = (X_{i1}^\star, \ldots, X_{id}^\star)$ is defined by

$$X_{ij}^\star = \begin{cases} X_{ij} & \text{if } j \in M \\ 0 & \text{otherwise,} \end{cases}$$

and $\mathbf{X}_i^{(n)} = \mathbf{X}_i^\star$ as soon as $M \subset M_i^{n+1-i}$.

We first prove the inclusion

$$\mathcal{A}_n \subset \left[|\{j \in \mathcal{L}_n : S(\mathbf{X}^\star, \mathbf{X}_j^\star) > 1/2\}| \leq k_n \right]. \tag{6.1}$$

Take $i \in \mathcal{L}_n^c$ such that $\mathbf{X}_i^{(n)}$ is among the k_n-MS of $\mathbf{X}^\star$ in $\{\mathbf{X}_i^{(n)}, i \in \mathcal{R}_n\}$. Then, for all $j \in \mathcal{L}_n$ such that $S(\mathbf{X}^\star, \mathbf{X}_j^\star) > 1/2$, we have

$$S(\mathbf{X}^\star, \mathbf{X}_j^\star) > \frac{1}{2} \geq p_i^{(n)} \bar{S}(\mathbf{X}^\star, \mathbf{X}_i^{(n)}) = S(\mathbf{X}^\star, \mathbf{X}_i^{(n)})$$

since $p_i^{(n)} \leq 1 - 1/|M| \leq 1/2$ if $|M| \geq 2$. If

$$|\{j \in \mathcal{L}_n : S(\mathbf{X}^\star, \mathbf{X}_j^\star) > 1/2\}| > k_n,$$

then $\mathbf{X}_i^{(n)}$ is not among the k_n-MS of $\mathbf{X}^\star$ among the $\{\mathbf{X}_i^{(n)}, i \in \mathcal{R}_n\}$. This contradicts the assumption on $\mathbf{X}_i^{(n)}$ and proves inclusion (6.1).

Next, define $\mathbf{Z}^\star = \mathbf{X}^\star / \|\mathbf{X}^\star\|$, $\mathbf{Z}_i^\star = \mathbf{X}_i^\star / \|\mathbf{X}_i^\star\|$, $i = 1, \ldots, n$, and let $\xi(\mathbf{Z}^\star) = \mathbb{P}(S(\mathbf{Z}^\star, \mathbf{Z}_1^\star) > 1/2 \,|\, \mathbf{Z}^\star)$. If $k_n - |\mathcal{L}_n|\xi(\mathbf{Z}^\star) \leq -(1/2)|\mathcal{L}_n|\xi(\mathbf{Z}^\star)$ and $\mathcal{L}_n \neq \emptyset$, we deduce from (6.1) that

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{A}_n \,\middle|\, \mathcal{L}_n, \mathbf{Z}^\star \right) \\
&\quad \leq \mathbb{P}\left(\sum_{j \in \mathcal{L}_n} \mathbf{1}_{\{S(\mathbf{Z}^\star, \mathbf{Z}_j^\star) > 1/2\}} \leq k_n \,\middle|\, \mathcal{L}_n, \mathbf{Z}^\star \right) \\
&\quad = \mathbb{P}\left(\sum_{j \in \mathcal{L}_n} \left(\mathbf{1}_{\{S(\mathbf{Z}^\star, \mathbf{Z}_j^\star) > 1/2\}} - \xi(\mathbf{Z}^\star) \right) \leq k_n - |\mathcal{L}_n|\xi(\mathbf{Z}^\star) \,\middle|\, \mathcal{L}_n, \mathbf{Z}^\star \right) \\
&\quad \leq \mathbb{P}\left(\sum_{j \in \mathcal{L}_n} \left(\mathbf{1}_{\{S(\mathbf{Z}^\star, \mathbf{Z}_j^\star) > 1/2\}} - \xi(\mathbf{Z}^\star) \right) \leq -\frac{1}{2}|\mathcal{L}_n|\xi(\mathbf{Z}^\star) \,\middle|\, \mathcal{L}_n, \mathbf{Z}^\star \right) \\
&\quad \leq \frac{4|\mathcal{L}_n|\xi(\mathbf{Z}^\star)}{\left(|\mathcal{L}_n|\xi(\mathbf{Z}^\star)\right)^2} = \frac{4}{|\mathcal{L}_n|\xi(\mathbf{Z}^\star)} \\
&\qquad \text{(by Tchebychev's inequality).}
\end{aligned}$$

In the last inequality, we use the fact that, since $\sigma(M) \subset \sigma(\mathbf{Z}^\star)$, the random variables $\{\mathbf{Z}_i^\star, i \in \mathcal{L}_n\}$ are independent conditionally on $\mathbf{Z}^\star$ and $\mathcal{L}_n$. Using again the inclusion $\sigma(M) \subset \sigma(\mathbf{Z}^\star)$, we obtain, on the event $[\mathcal{L}_n \neq 0]$,

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{A}_n \,\middle|\, \mathcal{L}_n, M\right) \\
&\quad = \mathbb{E}\left[\mathbb{P}\left(\mathcal{A}_n \middle| \mathcal{L}_n, \mathbf{Z}^\star\right) \middle| \mathcal{L}_n, M\right] \\
&\quad \leq \frac{4}{|\mathcal{L}_n|}\mathbb{E}\left[\frac{1}{\xi(\mathbf{Z}^\star)} \,\middle|\, \mathcal{L}_n, M\right] + \mathbb{P}\left(k_n - |\mathcal{L}_n|\xi(\mathbf{Z}^\star) > -\frac{1}{2}|\mathcal{L}_n|\xi(\mathbf{Z}^\star) \,\middle|\, \mathcal{L}_n, M\right) \\
&\quad = \frac{4}{|\mathcal{L}_n|}\mathbb{E}\left[\frac{1}{\xi(\mathbf{Z}^\star)} \,\middle|\, M\right] + \mathbb{P}\left(|\mathcal{L}_n|\xi(\mathbf{Z}^\star) < 2k_n \,\middle|\, \mathcal{L}_n, M\right).
\end{aligned}$$

Applying Lemma 6.3, on the event $[\mathcal{L}_n \neq \emptyset]$,

$$\begin{aligned}
&\mathbb{P}\left(\mathcal{A}_n \,\middle|\, \mathcal{L}_n, M\right) \\
&\leq \frac{4}{|\mathcal{L}_n|}\mathbb{E}\left[\frac{1}{\xi(\mathbf{Z}^\star)} \,\middle|\, M\right] + 2\mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|} \,\middle|\, \mathcal{L}_n\right]\mathbb{E}\left[\frac{1}{\xi(\mathbf{Z}^\star)} \,\middle|\, M\right] + \mathbb{E}\left[\frac{|\mathcal{L}_n^c|}{|\mathcal{R}_n|} \,\middle|\, \mathcal{L}_n, M\right].
\end{aligned}$$

Moreover, by fact 4.1,

$$\xi(\mathbf{Z}^\star) = \mathbb{P}\left(S(\mathbf{Z}^\star, \mathbf{Z}_1^\star) > \frac{1}{2} \,\middle|\, \mathbf{Z}^\star\right) \geq \mathbb{P}\left(d^2(\mathbf{Z}^\star, \mathbf{Z}_1^\star) \leq \frac{1}{2} \,\middle|\, \mathbf{Z}^\star\right).$$

Thus, denoting by ν^M the distribution of $\mathbf{Z}^\star$ conditionally to M, we deduce from Lemma 6.4 that

$$\mathbb{E}\left[\frac{1}{\xi(\mathbf{Z}^\star)} \,\middle|\, M\right] \leq \int \frac{1}{\nu^M(\mathcal{B}(\mathbf{z}, 1/\sqrt{2}))}\nu^M(\mathrm{d}\mathbf{z}) \leq C,$$

where the constant C does not depend on M. Putting all the pieces together, we obtain

$$\mathbb{P}(\mathcal{A}_n) \leq C\left\{\mathbb{E}\left[\frac{1}{|\mathcal{L}_n|}\mathbf{1}_{\{\mathcal{L}_n \neq \emptyset\}}\right] + \mathbb{E}\left[\frac{k_n}{|\mathcal{R}_n|}\right] + \mathbb{E}\left[\frac{|\mathcal{L}_n^c|}{|\mathcal{R}_n|}\right]\right\} + \mathbb{P}(\mathcal{L}_n = \emptyset).$$

We conclude the proof with Lemma 6.1 and Lemma 6.2.

□

In the sequel, we let $\mathbf{X}_{(1)}^\star, \ldots, \mathbf{X}_{(|\mathcal{L}_n|)}^\star$ be the sequence $\{\mathbf{X}_i^\star, i \in \mathcal{L}_n\}$ reordered according to decreasing similarities $S(\mathbf{X}^\star, \mathbf{X}_i^\star), i \in \mathcal{L}_n$, that is,

$$S\left(\mathbf{X}^\star, \mathbf{X}_{(1)}^\star\right) \geq \ldots \geq S\left(\mathbf{X}^\star, \mathbf{X}_{(|\mathcal{L}_n|)}^\star\right).$$

Lemma 6.6 below states the rate of convergence to 1 of $S(\mathbf{X}^\star, \mathbf{X}_{(1)}^\star)$.

Lemma 6.6 *Suppose that $|M| \geq 4$. Then there exists $C > 0$ such that, on the event $[\mathcal{L}_n \neq \emptyset]$,*

$$1 - \mathbb{E}\left[S(\mathbf{X}^\star, \mathbf{X}^\star_{(1)}) \,|\, M, \mathcal{L}_n\right] \leq \frac{C}{|\mathcal{L}_n|^{2/(|M|-1)}}.$$

Proof of Lemma 6.6. Observe that

$$\begin{aligned}
&\mathbb{E}\left[1 - S(\mathbf{X}^\star, \mathbf{X}^\star_{(1)}) \,|\, \mathbf{X}^\star, \mathcal{L}_n\right] \\
&\qquad = \int_0^1 \mathbb{P}\left(1 - S(\mathbf{X}^\star, \mathbf{X}^\star_{(1)}) > \varepsilon \,\middle|\, \mathbf{X}^\star, \mathcal{L}_n\right) \mathrm{d}\varepsilon \\
&\qquad = \int_0^1 \mathbb{P}\left(\forall i \in \mathcal{L}_n : 1 - S(\mathbf{X}^\star, \mathbf{X}^\star_i) > \varepsilon \,\middle|\, \mathbf{X}^\star, \mathcal{L}_n\right) \mathrm{d}\varepsilon.
\end{aligned}$$

Since $\sigma(M) \subset \sigma(\mathbf{X}^\star)$, given $\mathbf{X}^\star$ and $\mathcal{L}_n$, the random variables $\{\mathbf{X}^\star_i, i \in \mathcal{L}_n\}$ are independent and identically distributed. Hence,

$$\mathbb{E}\left[1 - S(\mathbf{X}^\star, \mathbf{X}^\star_{(1)}) \,|\, \mathbf{X}^\star, \mathcal{L}_n\right] = \int_0^1 \left[\mathbb{P}\left(1 - S(\mathbf{X}^\star, \mathbf{X}^\star_1) > \varepsilon \,\middle|\, \mathbf{X}^\star\right)\right]^{|\mathcal{L}_n|} \mathrm{d}\varepsilon.$$

Denote by ν^M the conditional distribution of $\mathbf{X}^\star/\|\mathbf{X}^\star\|$ given M. The support of ν^M is contained in both the unit sphere of $\mathbb{R}^d$ and in a $|M|$-dimensional vector space. Thus, for simplicity, we shall consider that the support of ν^M is contained in the unit sphere of $\mathbb{R}^{|M|}$. Let $\mathcal{B}^{|M|}(\mathbf{x}, r)$ be the closed Euclidean ball in $\mathbb{R}^{|M|}$ centered at $\mathbf{x}$ of radius r. Since $\mathbf{X}^\star$ (resp. $\mathbf{X}^\star_1$) only depends on M and $\mathbf{X}$ (resp. $\mathbf{X}_1$), then, given $\mathbf{X}^\star$, the random variable $\mathbf{X}^\star_1/\|\mathbf{X}^\star_1\|$ is distributed according to ν^M. Thus, for any $\varepsilon > 0$, we may write (fact 4.1)

$$\mathbb{P}\left(1 - S(\mathbf{X}^\star, \mathbf{X}^\star_1) > \varepsilon \,\middle|\, \mathbf{X}^\star\right) = 1 - \nu^M\left(\mathcal{B}^{|M|}\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}, \sqrt{2\varepsilon}\right)\right),$$

and, consequently,

$$\mathbb{E}\left[1 - S(\mathbf{X}^\star, \mathbf{X}^\star_{(1)}) \,|\, \mathbf{X}^\star, \mathcal{L}_n\right] = \int_0^1 \left[1 - \nu^M\left(\mathcal{B}^{|M|}\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}, \sqrt{2\varepsilon}\right)\right)\right]^{|\mathcal{L}_n|} \mathrm{d}\varepsilon.$$

Using the inclusion $\sigma(M) \subset \sigma(\mathbf{X}^\star)$, we obtain

$$\begin{aligned}
&\mathbb{E}\left[1 - S(\mathbf{X}^\star, \mathbf{X}^\star_{(1)}) \,|\, M, \mathcal{L}_n\right] \\
&\quad = \int_0^1 \mathbb{E}\left[\left\{1 - \nu^M\left(\mathcal{B}^{|M|}\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}, \sqrt{2\varepsilon}\right)\right)\right\}^{|\mathcal{L}_n|} \,\middle|\, M, \mathcal{L}_n\right] \mathrm{d}\varepsilon. \qquad (6.2)
\end{aligned}$$

Fix $\varepsilon > 0$, and denote by $\mathcal{S}(M)$ the support of ν^M. There exists Euclidean balls $A_1, \ldots, A_{N(\varepsilon)}$ in $\mathbb{R}^{|M|}$ with radius $\sqrt{2\varepsilon}/2$ such that

$$\mathcal{S}(M) \subset \bigcup_{j=1}^{N(\varepsilon)} A_j \quad \text{and} \quad N(\varepsilon) \leq \frac{C}{\varepsilon^{(|M|-1)/2}},$$

for some $C > 0$ which may be chosen independently of M. Clearly, if $\mathbf{x} \in A_j \cap \mathcal{S}(M)$, then $A_j \subset \mathcal{B}^{|M|}(\mathbf{x}, \sqrt{2\varepsilon})$. Thus,

$$\begin{aligned}
&\mathbb{E}\left[\left\{1 - \nu^M\left(\mathcal{B}^{|M|}\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}, \sqrt{2\varepsilon}\right)\right)\right\}^{|\mathcal{L}_n|} \Big| M, \mathcal{L}_n\right] \\
&\quad \leq \sum_{j=1}^{N(\varepsilon)} \int_{A_j} \mathbb{E}\left[\left\{1 - \nu^M\left(\mathcal{B}^M\left(\frac{\mathbf{X}^\star}{\|\mathbf{X}^\star\|}, \sqrt{2\varepsilon}\right)\right)\right\}^{|\mathcal{L}_n|} \Big| M, \mathcal{L}_n\right] \nu^M(\mathrm{d}\mathbf{x}) \\
&\quad \leq \sum_{j=1}^{N(\varepsilon)} \int_{A_j} \left(1 - \nu^M(A_j)\right)^{|\mathcal{L}_n|} \nu^M(\mathrm{d}\mathbf{x}) \\
&\quad \leq \sum_{j=1}^{N(\varepsilon)} \nu^M(A_j) \left(1 - \nu^M(A_j)\right)^{|\mathcal{L}_n|} \\
&\quad \leq N(\varepsilon) \max_{t \in [0,1]} t(1-t)^{|\mathcal{L}_n|} \\
&\quad \leq \frac{C}{|\mathcal{L}_n|\, \varepsilon^{(|M|-1)/2}}.
\end{aligned}$$

Combining this inequality and equality (6.2), we obtain

$$\mathbb{E}\left[1 - S(\mathbf{X}^\star, \mathbf{X}^\star_{(1)}) \,|\, M, \mathcal{L}_n\right] \leq \int_0^1 \min\left(1, \frac{C}{|\mathcal{L}_n|\, \varepsilon^{(|M|-1)/2}}\right) \mathrm{d}\varepsilon.$$

Since $|M| \geq 4$, an easy calculation shows that there exists $C > 0$ such that

$$\mathbb{E}\left[1 - S(\mathbf{X}^\star, \mathbf{X}^\star_{(1)}) \,|\, M, \mathcal{L}_n\right] \leq \frac{C}{|\mathcal{L}_n|^{2/(|M|-1)}},$$

which leads to the desired result.

□

Acknowledgments. The authors are greatly indebted to Albert Benveniste for pointing out this problem. They also thank Kevin Bleakley and Toby Hocking for their careful reading of the paper, and two referees and the Associate Editor for valuable comments and insightful suggestions.

References

[1] J. Abernethy, F.R. Bach, T. Evgeniou, and J.-P. Vert. A new approach to collaborative filtering: Operator estimation with spectral regularization. *J. Mach. Learn. Res.*, 10:803–826, 2009.

[2] G. Adomavicius, R. Sankaranarayanan, S. Sen, and A. Tuzhilin. Incorporating contextual information in recommender systems using a multidimensional approach. *ACM Trans. Info. Syst.*, 2005.

[3] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.*, 17:734–749, 2005.

[4] J.S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of 14th Conference on Uncertainty in Artificial Intelligence*, pages 43–52, 1998.

[5] E.J. Candès and Y. Plan. Matrix completion with noise. Preprint: http://www.acm.caltech.edu/~emmanuel/papers/NoisyCompletion.pdf.

[6] E.J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 2009. In press.

[7] S.H. Choi, S. Kang, and Y.J. Jeon. Personalized recommendation system based on product specification values. *Expert Systems with Applications*, 31:607–616, 2006.

[8] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, New-York, 1996.

[9] L. Györfi, M. Kohler, A. Krzyżak, and H. Walk. *A Distribution Free Theory of Nonparametric Regression*. Springer-Verlag, 2002.

[10] D. Heckerman, D.M. Chickering, C. Meek, R. Rounthwaite, and C. Kadie. Dependency networks for density estimation, collaborative filtering, and data visualization. *J. Mach. Learn. Res.*, 1:49–75, 2000.

[11] W.C. Hill, L. Stead, M. Rosenstein, and G.W. Furnas. Recommending and evaluating choices in a virtual community of use. In *Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems*, pages 194–201, 1995.

[12] M. Montaner, B. Lopez, and J.L.D. Rosa. A taxonomy of recommender agents on the Internet. *Artificial Intelligence Review*, 19:285–330, 2003.

[13] P. Resnick, N. Iakovou, M. Sushak, P. Bergstrom, and J. Riedl. Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 Computer Supported Cooperative Work Conference*, pages 175–186, 1994.

[14] R. Salakhutdinov, A. Mnih, and G. Hinton. Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th International Conference on Machine Learning*, pages 791–798, 2007.

[15] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International WWW Conference*, pages 285–295, 2001.

[16] U. Shardanand and P. Maes. Social information filtering: Algorithms for automating "word of mouth". In *Proceedings of the Conference on Human Factors in Computing Systems*, pages 210–217, 1995.