The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)

Clipped Matrix Completion: A Remedy for Ceiling Effects

Takeshi Teshima,[1,2] Miao Xu,[2] Issei Sato,[1,2] Masashi Sugiyama[1,2]
[1]The University of Tokyo
[2]RIKEN
teshima@ms.k.u-tokyo.ac.jp, miao.xu@riken.jp, {sato, sugi}@k.u-tokyo.ac.jp

Abstract

We consider the problem of recovering a low-rank matrix from its clipped observations. Clipping is conceivable in many scientific areas that obstructs statistical analyses. On the other hand, *matrix completion* (MC) methods can recover a low-rank matrix from various information deficits by using the principle of low-rank completion. However, the current theoretical guarantees for low-rank MC do not apply to clipped matrices, as the deficit depends on the underlying values. Therefore, the feasibility of *clipped matrix completion* (CMC) is not trivial. In this paper, we first provide a theoretical guarantee for the exact recovery of CMC by using a trace-norm minimization algorithm. Furthermore, we propose practical CMC algorithms by extending ordinary MC methods. Our extension is to use the squared *hinge* loss in place of the squared loss for reducing the penalty of over-estimation on clipped entries. We also propose a novel regularization term tailored for CMC. It is a combination of two trace-norm terms, and we theoretically bound the recovery error under the regularization. We demonstrate the effectiveness of the proposed methods through experiments using both synthetic and benchmark data for recommendation systems.

1 Introduction

Ceiling effect is a measurement limitation that occurs when the highest possible score on a measurement instrument is reached, thereby decreasing the likelihood that the instrument has accurately measured in the intended domain (Salkind 2010). In this paper[1], we investigate methods for restoring a matrix data from ceiling effects.

1.1 Ceiling effect

Ceiling effect has long been discussed across a wide range of scientific fields such as sociology (DeMaris 2004), educational science (Kaplan 1992; Benjamin 2005), biomedical research (Austin and Brunner 2003; Cox and Oakes 1984), and health science (Austin, Escobar, and Kopec 2000; Catherine et al. 2004; Voutilainen et al. 2016; Rodrigues et al. 2013), because it is a crucial information deficit known to inhibit effective statistical analyses (Austin and Brunner 2003).

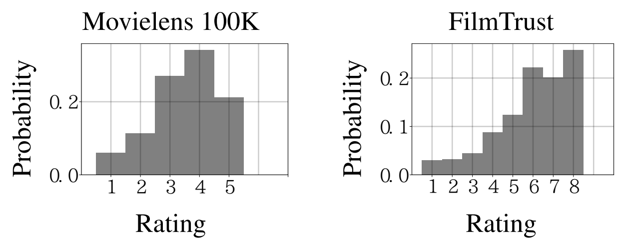


Figure 1: Ceiling effects may also exist in standard benchmark data sets of recommendation systems (details of the data are described in Section 6). Histograms of the rated values are plotted. The right-truncated look of the histogram is typical for a variable under ceiling effects (Greene 2012).

The ceiling effect is also conceivable in the context of machine learning, e.g., in recommendation systems with a five-star rating. After rating an item with a five-star, a user may find another item much better later. In this case, the true rating for the latter item should be above five, but the recorded value is still a five-star. As a matter of fact, we can observe right-truncated shapes indicating ceiling effects in the histograms of well-known benchmark data sets for recommendation systems, as shown in Figure 1.

Restoring data from ceiling effects can lead to benefits in many fields. For example, in biological experiments to measure the adenosine triphosphate (ATP) level, it is known that the current measurement method has a technical upper bound (Yaginuma et al. 2014). In such a case, by measuring multiple cells in multiple environments, we may recover the true ATP levels which can provide us with further findings. In the case of recommendation systems, we may be able to find latent superiority or inferiority between items with the highest ranking and recommend unobserved entries better.

In this paper, we investigate methods for restoring a matrix data from ceiling effects. In particular, we consider the recovery of a clipped matrix, i.e., elements of the matrix are clipped at a predefined threshold in advance of observation, because ceiling effects are often modeled as a clipping phenomenon (Austin and Brunner 2003).

[1]A longer version of this paper including Appendix is available at https://arxiv.org/abs/1809.04997.

4	7	4	7	4
0	3	6	15	12
4	6	2	2	0
2	6	7	16	12
8	13	6	9	4

(a) True matrix $\mathbf{M}$

	7	4	7	4
0	3	6	10	10
4	6	2	2	0
2	6	7	10	10
8	10	6	9	4

(b) Observed $\mathbf{M}^{\mathrm{c}}_{\Omega}$

4.0	7.0	4.0	7.0	4.0
-0.0	3.0	6.0	14.9	11.9
4.0	6.0	2.0	2.0	0.0
2.0	6.0	7.0	15.9	11.9
8.0	13.0	6.0	9.0	4.0

(c) Restored $\widehat{\mathbf{M}}$

Figure 2: Illustration of the task of CMC. The true low-rank matrix $\mathbf{M}$ has a distinct structure of large values. However, the observed data $\mathbf{M}^{\mathrm{c}}_{\Omega}$ is clipped at a predefined threshold $C = 10$. The goal of CMC is to restore $\mathbf{M}$ from the value of C and $\mathbf{M}^{\mathrm{c}}_{\Omega}$. The restored matrix $\widehat{\mathbf{M}}$ is an actual result of applying a proposed method (Fro-CMC).

1.2 Our problem: clipped matrix completion (CMC)

We consider the recovery of a low-rank matrix whose observations are clipped at a predefined threshold (Figure 2). We call this problem *clipped matrix completion* (CMC). Let us first introduce its background, low-rank *matrix completion*.

Low-rank *matrix completion* (MC) aims to recover a low-rank matrix from various information deficits, e.g., missing (Cands and Recht 2009; Recht 2011; Chen et al. 2015; Kirly, Theran, and Tomioka 2015), noise (Cands and Plan 2010), or discretization (Davenport et al. 2014; Lan, Studer, and Baraniuk 2014; Bhaskar 2016). The principle to enable low-rank MC is the dependency among entries of a low-rank matrix; each element can be expressed as the inner product of latent feature vectors of the corresponding row and column. With the principle of low-rank MC, we may be able to recover the entries of a matrix from a ceiling effect.

Clipped matrix completion (CMC). The CMC problem is illustrated in Figure 2. It is a problem to recover a low-rank matrix from random observations of its entries.

Formally, the goal of CMC in this paper can be stated as follows. Let $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ be the ground-truth low-rank matrix where $n_1, n_2 \in \mathbb{N}$, and $C \in \mathbb{R}$ be the clipping threshold. Let $\mathrm{Clip}(\cdot) := \min\{C, \cdot\}$ be the clipping operator that operates on matrices element-wise. We observe a random subset of entries of $\mathbf{M}^{\mathrm{c}} := \mathrm{Clip}(\mathbf{M})$. The set of observed indices is denoted by Ω. The goal of CMC is to accurately recover $\mathbf{M}$ from $\mathbf{M}^{\mathrm{c}}_{\Omega} := \{M^{\mathrm{c}}_{ij}\}_{(i,j)\in\Omega}$ and C.

Limitations of MC. There are two limitations regarding the application of existing MC methods to CMC.

1. The applicability of the principle of low-rank MC to clipped matrices is non-trivial because clipping occurs depending on the underlying values whereas the existing theoretical guarantees of MC methods presume the information deficit (e.g., missing or noise) to be independent of the values (Bhojanapalli and Jain 2014; Chen et al. 2015; Liu, Liu, and Yuan 2017).
2. Most of the existing MC methods fail to take ceiling effects into account, as they assume that the observed values are equal to or close to the true values (Cands and Recht 2009; Keshavan, Montanari, and Oh 2010), whereas clipped values may have a large gap from the original values.

The goal of this paper is to overcome these limitations and to propose low-rank completion methods suited for CMC.

1.3 Our contribution and approach

From the perspective of MC research, our contribution is three-fold.

1) We provide a theoretical analysis to establish the validity of the low-rank principle in CMC (Section 2). To do so, we provide an exact recovery guarantee: a sufficient condition for a trace-norm minimization algorithm to perfectly recover the ground truth matrix with high probability. Our analysis is based on the notion of *incoherence* (Cands and Recht 2009; Recht 2011; Chen et al. 2015).

2) We propose practical algorithms for CMC (Section 3) and provide an analysis of the recovery error (Section 4). We propose practical CMC methods which are extensions of the Frobenius norm minimization that is frequently used for MC (Toh and Yun 2010). The simple idea of extension is to replace the squared loss function with the squared hinge loss to cancel the penalty of over-estimation on clipped entries. We also propose a regularizer consisting of two trace-norm terms, which is motivated by a theoretical analysis of a recovery error bound.

3) We conducted experiments using synthetic and real-world data to demonstrate the validity of the proposed methods (Section 6). Using synthetic data with known ground truth, we confirmed that the proposed CMC methods can actually recover randomly-generated matrices from clipping. We also investigated the improved robustness of CMC methods to the existence of clipped entries in comparison with ordinary MC methods. Using real-world data, we conducted two experiments to validate the effectiveness of the proposed CMC methods.

1.4 Additional notation

The symbols $\mathbf{M}, \mathbf{M}^{\mathrm{c}}, \mathbf{M}^{\mathrm{c}}_{\Omega}, \Omega, C$, and Clip are used throughout the paper. Let r be the rank of $\mathbf{M}$. The set of observed clipped indices is $\mathcal{C} := \{(i,j) \in \Omega : M^{\mathrm{c}}_{ij} = C\}$. Given a set of indices $\mathcal{S}$, we define its projection operator $\mathcal{P}_{\mathcal{S}} : \mathbb{R}^{n_1 \times n_2} \to \mathbb{R}^{n_1 \times n_2}$ by $(\mathcal{P}_{\mathcal{S}}(\mathbf{X}))_{ij} := \mathbb{1}\{(i,j) \in \mathcal{S}\}X_{ij}$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function giving 1 if the condition is true and 0 otherwise. We use $\|\cdot\|, \|\cdot\|_{\mathrm{tr}}, \|\cdot\|_{\mathrm{op}}, \|\cdot\|_{\mathrm{F}}$, and $\|\cdot\|_{\infty}$ for the Euclidean norm of vectors, the trace-norm, the operator norm, the Frobenius norm, the infinity norm of matrices, respectively. We also use $(\cdot)^{\top}$ for the transpose and define $[n] := \{1, 2, \ldots, n\}$ for $n \in \mathbb{N}$. For a notation table, please see Table 4 in Appendix.

2 Feasibility of the CMC problem

As noted earlier, it is not trivial if the principle of low-rank MC guarantees the recover of clipped matrices. In this section, we establish that the principle of low-rank completion is still valid for some matrices by providing a sufficient condition under which an exact recovery by trace-norm minimization is achieved with high probability.

We consider a *trace-norm minimization* for CMC:

$$\widehat{\mathbf{M}} \in \mathop{\arg\min}_{\mathbf{X}\in\mathbb{R}^{n_1\times n_2}} \|\mathbf{X}\|_{\mathrm{tr}} \text{ s.t. } \begin{cases}\mathcal{P}_{\Omega\setminus\mathcal{C}}(\mathbf{X}) = \mathcal{P}_{\Omega\setminus\mathcal{C}}(\mathbf{M}^{\mathrm{c}}),\\ \mathcal{P}_{\mathcal{C}}(\mathbf{M}^{\mathrm{c}}) \le \mathcal{P}_{\mathcal{C}}(\mathbf{X}),\end{cases} \tag{1}$$

where "s.t." stands for "subject to." Note that the optimization problem Eq. (1) is convex, and there are algorithms that can solve it numerically (Liu and Vandenberghe 2010).

2.1 Definitions and intuition of the information loss measures

Here, we define the quantities required for stating the theorem. The quantities reflect the difficulty of recovering $\mathbf{M}$, therefore the sufficient condition stated in the theorem will be that these quantities are small enough. Let us begin with the definition of coherence that captures how much the row and column spaces of a matrix is aligned with the standard basis vectors (Cands and Recht 2009; Recht 2011; Chen et al. 2015).

Def. 1 (Coherence and joint coherence (Chen et al. 2015))**.** *Let $\mathbf{X} \in \mathbb{R}^{n_1\times n_2}$ have a skinny singular value decomposition $\mathbf{X} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^\top$. We define*

$$\mu^{\mathrm{U}}(\mathbf{X}) := \max_{i\in[n_1]} \|\tilde{\mathbf{U}}_{i,\cdot}\|^2,\ \ \mu^{\mathrm{V}}(\mathbf{X}) := \max_{j\in[n_2]} \|\tilde{\mathbf{V}}_{j,\cdot}\|^2,$$

where $\tilde{\mathbf{U}}_{i,\cdot}$ ($\tilde{\mathbf{V}}_{j,\cdot}$) is the i-th (resp. j-th) row of $\tilde{\mathbf{U}}$ (resp. $\tilde{\mathbf{V}}$). Now the coherence of $\mathbf{M}$ is defined by

$$\mu_0 := \max\left\{\frac{n_1}{r}\mu^{\mathrm{U}}(\mathbf{M}), \frac{n_2}{r}\mu^{\mathrm{V}}(\mathbf{M})\right\}.$$

In addition, we define the following joint coherence:

$$\mu_1 := \sqrt{\frac{n_1 n_2}{r}}\|\mathbf{U}\mathbf{V}^\top\|_\infty.$$

The feasibility of CMC depends upon the amount of information that clipping can hide. To characterize the amount of information obtained from observations of $\mathbf{M}$, we define a subspace T that is also used in the existing recovery guarantees for MC (Cands and Recht 2009).

Def. 2 (The information subspace of $\mathbf{M}$ (Cands and Recht 2009))**.** *Let $\mathbf{M} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be a skinny singular value decomposition ($\mathbf{U} \in \mathbb{R}^{n_1\times r}, \mathbf{\Sigma} \in \mathbb{R}^{r\times r}$ and $\mathbf{V} \in \mathbb{R}^{n_2\times r}$). We define*

$$T := \mathrm{span}\big(\{\boldsymbol{u}_k\boldsymbol{y}^\top : k\in[r], \boldsymbol{y}\in\mathbb{R}^{n_1}\} \cup \{\boldsymbol{x}\boldsymbol{v}_k^\top : k\in[r], \boldsymbol{x}\in\mathbb{R}^{n_2}\}\big),$$

where $\boldsymbol{u}_k, \boldsymbol{v}_k$ are the k-th column of $\mathbf{U}$ and $\mathbf{V}$, respectively. Let $\mathcal{P}_T$ and $\mathcal{P}_{T^\perp}$ denote the projections onto T and $T^\perp$, respectively, where $\perp$ denotes the orthogonal complement.

Using T, we define the quantities to capture the amount of information loss due to clipping, in terms of different matrix norms representing different types of dependencies. To express the factor of clipping, we define a transformation $\mathcal{P}^*$ on $\mathbb{R}^{n_1\times n_2}$ that describes the amount of information left after clipping. Therefore, if these quantities are small, enough information for recovering $\mathbf{M}$ may be preserved after clipping.

Def. 3 (The information loss measured in various norms)**.** *Define*

$$\rho_{\mathrm{F}} := \sup_{\mathbf{Z}\in T\setminus\{\mathbf{O}\}:\|\mathbf{Z}\|_{\mathrm{F}}\le\|\mathbf{U}\mathbf{V}^\top\|_{\mathrm{F}}} \frac{\|\mathcal{P}_T\mathcal{P}^*(\mathbf{Z})-\mathbf{Z}\|_{\mathrm{F}}}{\|\mathbf{Z}\|_{\mathrm{F}}},$$

$$\rho_\infty := \sup_{\mathbf{Z}\in T\setminus\{\mathbf{O}\}:\|\mathbf{Z}\|_\infty\le\|\mathbf{U}\mathbf{V}^\top\|_\infty} \frac{\|\mathcal{P}_T\mathcal{P}^*(\mathbf{Z})-\mathbf{Z}\|_\infty}{\|\mathbf{Z}\|_\infty},$$

$$\rho_{\mathrm{op}} := \sqrt{r}\mu_1\left(\sup_{\substack{\mathbf{Z}\in T\setminus\{\mathbf{O}\}:\\ \|\mathbf{Z}\|_{\mathrm{op}}\le\sqrt{n_1n_2}\|\mathbf{U}\mathbf{V}^\top\|_{\mathrm{op}}}} \frac{\|\mathcal{P}^*(\mathbf{Z})-\mathbf{Z}\|_{\mathrm{op}}}{\|\mathbf{Z}\|_{\mathrm{op}}}\right),$$

where the operator $\mathcal{P}^* : \mathbb{R}^{n_1\times n_2}\to\mathbb{R}^{n_1\times n_2}$ is defined by

$$(\mathcal{P}^*(\mathbf{Z}))_{ij} = \begin{cases} Z_{ij} & \text{if } M_{ij} < C,\\ \max\{Z_{ij}, 0\} & \text{if } M_{ij} = C,\\ 0 & \text{otherwise}.\end{cases}$$

In addition, we define the following quantity that captures how much information of T depends on the clipped entries of $\mathbf{M}^{\mathrm{c}}$. If this quantity is small, enough information of T may be left in non-clipped entries.

Def. 4 (The importance of clipped entries for T)**.** *Define*

$$\nu_{\mathcal{B}} := \|\mathcal{P}_T\mathcal{P}_{\mathcal{B}}\mathcal{P}_T - \mathcal{P}_T\|_{\mathrm{op}},$$

where $\mathcal{B} := \{(i,j) : M_{ij} < C\}$.

We follow Chen et al. (2015) to assume the following observation scheme. As a result, it amounts to assuming that Ω is a result of random sampling where each entry is observed with probability p independently.

Assumption 1 (Assumption on the observation scheme)**.** *Let $p \in [0,1]$. Let $k_0 := \lceil\log_2(2\sqrt{2}\sqrt{n_1n_2r})\rceil$ and $q := 1-(1-p)^{1/k_0}$. For each $k = 1,\dots,k_0$, let $\Omega_k \subset [n_1]\times[n_2]$ be a random set of matrix indices that were sampled according to $\mathbb{P}((i,j)\in\Omega_k) = q$ independently. Then, Ω was generated by $\Omega = \bigcup_{k=1}^{k_0}\Omega_k$.*

The need for Assumption 1 is technical (Chen et al. 2015). Refer to the proof in Appendix D for details.

2.2 The theorem

We are now ready to state the theorem.

Theorem 1 (Exact recovery guarantee for CMC)**.** *Assume $\rho_{\mathrm{F}} < \frac{1}{2}, \rho_{\mathrm{op}} < \frac{1}{4}, \rho_\infty < \frac{1}{2}, \nu_{\mathcal{B}} < \frac{1}{2}$, and Assumption 1 for some $p \in [0,1]$. For simplicity of the statement, assume $n_1, n_2 \ge 2$ and $p \ge \frac{1}{n_1n_2}$. If, additionally,*

$$p \ge \min\left\{1, c_\rho \max({\mu_1}^2, \mu_0) r f(n_1, n_2)\right\}$$

is satisfied, then the solution of Eq. (1) is unique and equal to $\mathbf{M}$ *with probability at least* $1-\delta$, *where*

$$\begin{aligned} c_\rho &= \max\left\{\frac{24}{(1/2-\rho_{\mathrm{F}})^2}, \frac{8}{(1/4-\rho_{\mathrm{op}})^2}, \right. \\ &\qquad \left. \frac{8}{(1/2-\rho_\infty)^2}, \frac{8}{(1/2-\nu_{\mathcal{B}})^2}\right\}, \\ f(n_1, n_2) &= \mathcal{O}\left(\frac{(n_1+n_2)(\log(n_1 n_2))^2}{n_1 n_2}\right), \\ \delta &= \mathcal{O}\left(\frac{\log(n_1 n_2)}{(n_1+n_2)^2}\right). \end{aligned}$$

The proof and the precise expressions of f and δ are available in Appendix D. A more general form of Theorem 1 allowing for clipping from below is also available in Appendix E. The information losses (Def. 3 and Def. 4) appear neither in the order of p nor that of δ, but they appear as coefficients and deterministic conditions. The existence of such a deterministic condition is in accordance with the intuition that an all-clipped matrix can never be completed no matter how many entries are observed.

Note that $p > 1/(n_1 n_2)$ can be safely assumed when there is at least one observation. An intuition regarding the conditions on $\rho_{\mathrm{F}}, \rho_{\mathrm{op}}, \rho_\infty$, and $\nu_{\mathcal{B}}$ is that the singular vectors of $\mathbf{M}$ should not be too aligned with the clipped entries for the recovery to be possible, similarly to the intuition for the incoherence condition in previous theoretical works such as Cands and Recht (2009).

3 Practical algorithms

In this section, we introduce practical algorithms for CMC. The trace-norm minimization (Eq. (1)) is known to require impractical running time as the problem size increases from small to moderate or large (Cai, Cands, and Shen 2010).

A popular method for matrix completion is to minimize the squared error between the prediction and the observed value under some regularization (Toh and Yun 2010). We develop our CMC methods following this approach.

Throughout this section, $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ generally denotes an optimization variable, which may be further parametrized by $\mathbf{X} = \mathbf{P}\mathbf{Q}^\top$ (where $\mathbf{P} \in \mathbb{R}^{n_1 \times k}, \mathbf{Q} \in \mathbb{R}^{n_2 \times k}$ for some $k \leq \min(n_1, n_2)$). Regularization terms are denoted by $\mathcal{R}$, and regularization coefficients by $\lambda, \lambda_1, \lambda_2 \geq 0$.

Frobenius norm minimization for MC. In the MC methods based on the Frobenius norm minimization (Toh and Yun 2010), we define

$$f^{\mathrm{MC}}(\mathbf{X}) := \frac{1}{2}\|\mathcal{P}_\Omega(\mathbf{M}^{\mathrm{c}} - \mathbf{X})\|_{\mathrm{F}}^2, \tag{2}$$

and obtain the estimator by

$$\widehat{\mathbf{M}} \in \mathop{\arg\min}_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} f^{\mathrm{MC}}(\mathbf{X}) + \mathcal{R}(\mathbf{X}). \tag{3}$$

The problem in using this method for CMC is that it is not robust to clipped entries as the loss function is designed under the belief that the true values are close to the observed values. We extend this method for CMC with a simple idea.

The general idea of extension. The general idea of extension is *not to penalize* the estimator on clipped entries when the predicted value exceeds the observed value. Therefore, we modify the loss function to

$$f^{\mathrm{CMC}}(\mathbf{X}) = \frac{1}{2}\|\mathcal{P}_{\Omega \setminus \mathcal{C}}(\mathbf{M}^{\mathrm{c}} - \mathbf{X})\|_{\mathrm{F}}^2 + \frac{1}{2}\sum_{(i,j) \in \mathcal{C}} (\mathbf{M}^{\mathrm{c}}_{ij} - \mathbf{X}_{ij})_+^2, \tag{4}$$

where $(\cdot)_+^2 := (\max(0, \cdot))^2$ is the squared hinge loss, which does not penalize over-estimation. Then we obtain the estimator by

$$\widehat{\mathbf{M}} \in \mathop{\arg\min}_{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}} f^{\mathrm{CMC}}(\mathbf{X}) + \mathcal{R}(\mathbf{X}). \tag{5}$$

From here, we discuss three designs of regularization terms for CMC. The methods are summarized in Table 1, and further details of the algorithms can be found in Appendix A.

Double trace-norm regularization. We first propose to use $\mathcal{R}(\mathbf{X}) = \lambda_1\|\mathbf{X}\|_{\mathrm{tr}} + \lambda_2\|\mathrm{Clip}(\mathbf{X})\|_{\mathrm{tr}}$. For this method, we will conduct a theoretical analysis of the recovery error in Section 4. For optimization, we employ an iterative method based on subgradient descent (Avron et al. 2012). Even though the second term, $\lambda_2\|\mathrm{Clip}(\mathbf{X})\|_{\mathrm{tr}}$, is a composition of a nonlinear mapping and a non-smooth convex function, we can take advantage of its simple structure to approximate it with a convex function of $\mathbf{X}$ whose subgradient can be calculated for each iteration. We refer to this algorithm as *DTr-CMC* (Double Trace-norm regularized CMC).

Trace-norm regularization. With trace-norm regularization $\mathcal{R}(\mathbf{X}) := \lambda\|\mathbf{X}\|_{\mathrm{tr}}$, the optimization problem Eq. (5) is a relaxation of the trace-norm minimization (Eq. (1)) by replacing the exact constraints with the quadratic penalties (Eq. (2) for MC and Eq. (4) for CMC). For optimization, we employ an accelerated proximal gradient (APG) algorithm proposed by Toh and Yun (2010), by taking advantage of the differentiability of the squared hinge loss. We refer to this algorithm as *Tr-CMC* (Trace-norm-regularized CMC), in contrast to *Tr-MC* (its MC counterpart; Toh and Yun 2010).

Frobenius norm regularization. This method first parametrizes $\mathbf{X}$ as $\mathbf{P}\mathbf{Q}^\top$ and use $\mathcal{R}(\mathbf{P}, \mathbf{Q}) := \lambda_1\|\mathbf{P}\|_{\mathrm{F}}^2 + \lambda_2\|\mathbf{Q}\|_{\mathrm{F}}^2$ for regularization. A commonly used method for optimization in the case of MC is the alternating least squares (ALS) method (Jain, Netrapalli, and Sanghavi 2013). Here, we employ an approximate optimization scheme motivated by ALS in our experiments. We refer to this algorithm as *Fro-CMC* (Frobenius-norm-regularized CMC), in contrast to *Fro-MC* (its MC counterpart; Jain, Netrapalli, and Sanghavi 2013).

4 Theoretical analysis for DTr-CMC

In this section, we provide a theoretical guarantee for DTr-CMC. Let G be the hypothesis space defined by

$$\begin{aligned} G = \big\{\mathbf{X} \in \mathbb{R}^{n_1 \times n_2} : &\|\mathbf{X}\|_{\mathrm{tr}}^2 \leq \beta_1\sqrt{k n_1 n_2}, \\ &\|\mathrm{Clip}(\mathbf{X})\|_{\mathrm{tr}}^2 \leq \beta_2\sqrt{k n_1 n_2}\big\} \end{aligned}$$

Table 1: List of the proposed methods for CMC (Fro: Frobenius norm, Tr: Trace-norm, Sq.hinge: Squared hinge loss, SUGD: SUb-Gradient Descent, APG: Accelerated Proximal Gradient, ALS: Alternating Least Squares, Param.: Parametrization, Reg.: Regularization, Opt.: Optimization).

Method	Param.	Loss on $\mathcal{C}$	Reg.	Opt.
DTr-CMC	$\mathbf{X}$	Sq. hinge	Tr + Tr	SUGD
Tr-CMC	$\mathbf{X}$	Sq. hinge	Tr	APG
Fro-CMC	$\mathbf{PQ}^\top$	Sq. hinge	Fro	ALS

for some $k \leq \min(n_1, n_2)$ and $\beta_1, \beta_2 \geq 0$. Here, we analyze the estimator

$$\widehat{\mathbf{M}} \in \underset{\mathbf{X} \in G}{\arg\min} \sum_{(i,j) \in \Omega} (M^{\mathrm{c}}_{ij} - \mathrm{Clip}(X_{ij}))^2. \tag{6}$$

The minimization objective of Eq. (6) is not convex. However, it is upper-bounded by the convex loss function f^{CMC} (Eq. (4)). The proof is provided in Appendix A.1. Therefore, DTr-CMC can be seen as a convex relaxation of Eq. (6) with constraints turned into regularization terms. To state our theorem, we define the unnormalized coherence of a matrix.

Def. 5 (Unnormalized coherence)**.** *Let $\mu(\mathbf{X})$ be unnormalized coherence defined by*

$$\mu(\mathbf{X}) = \max\{\mu^{\mathrm{U}}(\mathbf{X}), \mu^{\mathrm{V}}(\mathbf{X})\},$$

using μ^{U} and μ^{V} from Def. 1.

Now we are ready to state our theorem.

Theorem 2 (Theoretical guarantee for DTr-CMC)**.** *Suppose that $\mathbf{M} \in G$, and that Ω is generated by independent observation of entries with probability $p \in [0, 1]$. Let $\mu_G = \sup_{\mathbf{X} \in G} \mu(\mathrm{Clip}(\mathbf{X}))$, and $\widehat{\mathbf{M}}$ be a solution to the optimization problem Eq. (6). Then there exist universal constants C_0 and C_1, such that with probability at least $1 - C_1/(n_1 + n_2)$ we have*

$$\begin{aligned} &\sqrt{\frac{1}{n_1 n_2}\|\widehat{\mathbf{M}} - \mathbf{M}\|_{\mathrm{F}}^2} \\ &\leq \underbrace{\frac{\|\mathbf{M} - \mathbf{M}^{\mathrm{c}}\|_{\mathrm{F}}}{\sqrt{n_1 n_2}}}_{=B_1:\text{Complexity of data}} + \underbrace{\frac{\|\widehat{\mathbf{M}} - \mathrm{Clip}(\widehat{\mathbf{M}})\|_{\mathrm{F}}}{\sqrt{n_1 n_2}}}_{=B_2:\text{Complexity of hypothesis}} \\ &\quad + \underbrace{\frac{\|\mathrm{Clip}(\widehat{\mathbf{M}}) - \mathrm{Clip}(\mathbf{M})\|_{\mathrm{F}}}{\sqrt{n_1 n_2}}}_{=B_3:\text{Estimation error}}, \end{aligned} \tag{7}$$

and

$$\begin{aligned} &B_1 \leq (\sqrt{\beta_1} + \sqrt{\beta_2}) k^{\frac{1}{4}} (n_1 n_2)^{-\frac{1}{4}}, \\ &B_2 \leq (\sqrt{\beta_1} + \sqrt{\beta_2}) k^{\frac{1}{4}} (n_1 n_2)^{-\frac{1}{4}}, \\ &B_3 \leq \sqrt{C_0 \frac{2{\mu_G}^2 \beta_2}{p} \left(\frac{pk(n_1 + n_2) + k\log(n_1 + n_2)}{n_1 n_2}\right)^{\frac{1}{4}}}. \end{aligned}$$

We provide the proof in Appendix F. The right-hand side of Eq. (7) converges to zero as $n_1, n_2 \to \infty$ with p, k, β_1, and β_2 fixed. From this theorem, it is expected that if $\|\mathbf{M}\|_{\mathrm{tr}}$ and $\|\mathbf{M}^{\mathrm{c}}\|_{\mathrm{tr}}$ are believed to be small, DTr-CMC can accurately recover $\mathbf{M}$.

5 Related work

In this section, we describe related work from the literature on matrix completion and that on ceiling effects. Table 2 briefly summarizes the related work on matrix completion.

5.1 Matrix completion methods.

Theory. Our feasibility analysis in Section 2 followed the approach of Recht (2011) while some details of the proof were based on Chen et al. (2015). There is further research to weaken the assumption of the uniformly random observation (Bhojanapalli and Jain 2014). It may be relatively easy to incorporate such devices into our theoretical analysis.

Our theoretical analysis for DTr-CMC in Section 4 is inspired by the theory for 1-bit matrix completion (Davenport et al. 2014). The difference is that our theory effectively exploits the additional low-rank structure in the clipped matrix in addition to the original matrix.

Problem setting. Our problem setting of clipping can be related to quantized matrix completion (Q-MC; Lan, Studer, and Baraniuk 2014; Bhaskar 2016). Lan, Studer, and Baraniuk (2014) and Bhaskar (2016) formulated a probabilistic model which assigns discrete values according to a distribution conditioned on the underlying values of a matrix. Bhaskar (2016) provided an error bound for restoring the underlying values, assuming that the quantization model is fully known. The model of Q-MC can provide a different formulation for ceiling effects from ours by assuming the existence of latent random variables. However, Q-MC methods require the data to be fully discrete (Lan, Studer, and Baraniuk 2014; Bhaskar 2016). Therefore, neither their methods nor theories can be applied to real-valued observations. On the other hand, our methods and theories allow observations to be real-valued. The ceiling effect is worth studying independently of quantization, since the data analyzed under ceiling effects are not necessarily discrete.

Methodology. The use of the Frobenius norm for MC has been studied for MC from noisy data (Cands and Plan 2010; Toh and Yun 2010). Our algorithms are based on this line of research, while extending it for CMC.

Methodologically, Mareek, Richtrik, and Tak (2017) is closely related to our Fro-CMC. They considered completion of missing entries under "interval uncertainty" which yields interval constraints indicating the ranges in which the true values should reside. They employed the squared hinge loss for enforcing the interval constraints in their formulation, hence coinciding with our formulation of Fro-CMC. There are a few key differences between their work and ours. First, our motivations are quite different as we are analyzing a different problem from theirs. They considered completion of missing entries with robustness to uncertainty, whereas we considered recovery of clipped entries. Secondly, they did not provide any theoretical analysis of the

Table 2: Our target problem is the restoration of a low-rank matrix from clipping at a predefined threshold. No existing work has considered this type of information deficit.

Type of deficit	Related work
Missing	(Cands and Recht 2009) etc.
Noise	(Cands and Plan 2010) etc.
Quantization	(Bhaskar 2016) etc.
Clipping	This paper

problem. We provided an analysis by specifically investigating the problem of clipping. Lastly, as a minor difference, we employed an ALS-like algorithm whereas they used a coordinate descent method (Mareek, Richtrik, and Tak 2017; Marecek et al. 2018), as we found the ALS-like method to work well for moderately sized matrices.

5.2 Related work on ceiling effects

From the perspective of dealing with ceiling effects, the present paper adds a potentially effective method to the analysis of data affected by a ceiling effect. Ceiling effect is also referred to as *censoring* (Greene 2012) or *limited response variables* (DeMaris 2004). In this paper, we use "ceiling effect" to represent these phenomena. In *econometrics*, Tobit models are used to deal with ceiling effects (Greene 2012). In Tobit models, a *censored likelihood* is modeled and maximized with respect to the parameters of interest. Although this method is justified by the theory of M-estimation (Schnedler 2005; Greene 2012), its use for matrix completion is not justified. In addition, Tobit models require strong distributional assumptions, which is problematic especially if the distribution cannot be safely assumed.

6 Experimental results

In this section, we show the results of experiments to compare the proposed CMC methods to the MC methods.

6.1 Experiment with synthetic data

We conducted an experiment to recover randomly generated data from clipping. The primary purpose of the experiment was to confirm that the principle of low-rank completion is still effective for the recovery of a clipped matrix, as indicated by Theorem 1. Additionally, with the same experiment, we investigated how sensitive the MC methods are to the clipped entries by looking at the growth of the recovery error in relation to increased rates of clipping.

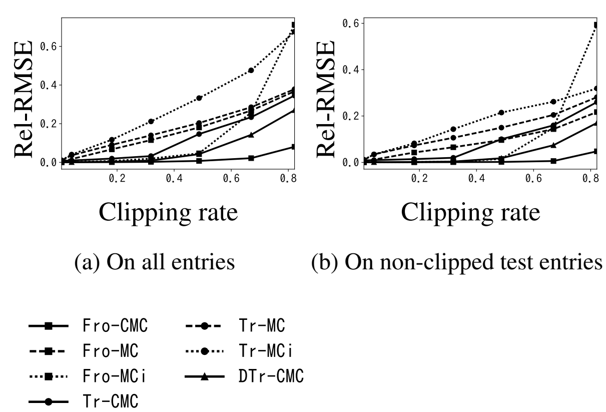
Data generation process. We randomly generated non-negative integer matrices of size 500×800 that are exactly rank-30 with the fixed magnitude parameter $L = 15$ (see Appendix B). The generated elements of matrix $\mathbf{M}$ were randomly split into three parts with ratio $(0.8, 0.1, 0.1)$. Then the first part was clipped at the threshold C (varied over $\{5, 6, 7, 8, 9, 11, 13\}$) to generate the training matrix $\mathbf{M}^{\mathrm{c}}_{\Omega}$ (therefore, $p = 0.8$). The remaining two parts (without thresholding) were treated as the validation ($\mathbf{M}^{\mathrm{v}}$) and testing ($\mathbf{M}^{\mathrm{t}}$) matrices, respectively.

Evaluation metrics. We used the relative root mean square error (rel-RMSE) as the evaluation metric, and we considered a result as a good recovery if the error is of order 10^{-2} (Toh and Yun 2010). We separately reported the rel-RMSE on two sets of indices: all the indices of $\mathbf{M}$, and the test entries whose true values are below the clipping threshold. For hyperparameter tuning, we used the rel-RMSE after clipping on validation indices: $\frac{\|\mathrm{Clip}(\widehat{\mathbf{M}})-\mathrm{Clip}(\mathbf{M}^{\mathrm{v}})\|_{\mathrm{F}}}{\|\mathrm{Clip}(\mathbf{M}^{\mathrm{v}})\|_{\mathrm{F}}}$. We reported the mean of five independent runs. The clipping rate was calculated by the ratio of entries of $\mathbf{M}$ above C.

Compared methods. We evaluated the proposed methods (DTr-CMC, Tr-CMC, and Fro-CMC) and their MC counterparts (Tr-MC and Fro-MC). We also applied MC methods after ignoring all clipped training entries (Tr-MCi and Fro-MCi, with "i" standing for "ignore"). While this treatment wastes some data, it may improve the robustness of MC methods to the existence of clipped entries.

Result 1: The validity of low-rank completion. In Figure (3a), we show the rel-RMSE for different clipping rates. The proposed methods successfully recover the true matrices with very low error of order 10^{-2} even when half of the observed training entries are clipped. One of them (Fro-CMC) is able to successfully recover the matrix after the clipping rate was above 0.6. This may be explained in part by the fact that the synthetic data were exactly low rank, and that the correct rank was in the search space of the bilinear model of the Frobenius norm based methods.

Result 2: The robustness to the existence of clipped training entries. In Figure (3b), the recovery error of MC methods on non-clipped entries increased with the rate of clipping. This indicates the disturbance effect of the clipped entries for ordinary MC methods. The MC methods with the clipped entries ignored (Tr-MCi and Fro-MCi) were also prone to increasing test error on non-clipped entries for high clipping rates, most likely due to wasting too much infor-



(a) On all entries (b) On non-clipped test entries

Figure 3: Relative RMSE for varied C (Dotted: previous MC methods, Solid: proposed CMC methods).

mation. On the other hand, the proposed methods show improved profiles of growth, indicating improved robustness.

6.2 Experiments with real-world data

We conducted two experiments using real-world data. The difficulty of evaluating CMC with real-world data is that there are no known true values unaffected by the ceiling effect. Therefore, instead of evaluating the accuracy of recovery, we evaluated the performance of distinguishing entries with the ceiling effect and those without. We considered two binary classification tasks in which we predict whether held-out test entries are of high values. The tasks would be reasonable because the purpose of a recommendation system is usually to predict which entries have *high* scores.

Preparation of data sets. We used the following benchmark data sets of recommendation systems.

- FilmTrust (Guo, Zhang, and Yorke-Smith 2013)[2] consists of ratings obtained from 1,508 users to 2,071 movies on a scale from 0.5 to 4.0 with a stride of 0.5 (approximately 99.0% missing). For ease of comparison, we doubled the ratings so that they are integers from 1 to 8.
- Movielens (100K)[3] consists of ratings obtained from 943 users to 1,682 movies on an integer scale from 1 to 5 (approximately 94.8% missing).

Task 1: Using artificially clipped training data. We artificially clipped the training data at threshold C and predicted whether the test entries were originally above C. We used $C = 7$ for FilmTrust and $C = 4$ for Movielens. For testing, we made positive prediction for entries above $C + 0.5$ and negative prediction otherwise.

Task 2: Using raw data. We used the raw training data and predicted whether the test entries are equal to the maximum value of the rating scale (i.e., the underlying values are at least the maximum value). For CMC methods, we set C to the maximum value, i.e., $C = 8$ for FilmTrust and $C = 5$ for Movielens. For testing, we made positive prediction for entries above $C - 0.5$ and negative prediction otherwise.

Protocols and evaluation metrics. In both experiments, we first split the observed entries into three groups with ratio $(0.8, 0.1, 0.1)$, which were used as training, validation, and test entries. Then for the first task, we artificially clipped the training data at C. If a user or an item had no training entries, we removed them from all matrices.

We measured the performance by the f_1 score. Hyperparameters were selected by the f_1 score on the validation entries. We reported the mean and the standard error after five independent runs.

Compared methods. We compared the proposed CMC methods with the corresponding MC methods. The baseline (indicated as "baseline") is to make prediction positively for all entries, for which the recall is 1 and the precision is the ratio of the positive data. This is the best baseline in terms of the f_1 score without looking at data.

[2] https://www.librec.net/datasets.html

[3] http://grouplens.org/datasets/movielens/100k/

Table 3: Results of the two tasks measured in f_1. Bold-face indicates the highest score.

Data	Methods	Task 1 f_1	Task 2 f_1
Film Trust	DTr-CMC	**0.47 (0.01)**	**0.46 (0.01)**
	Fro-CMC	0.35 (0.01)	0.40 (0.01)
	Fro-MC	0.27 (0.01)	0.35 (0.01)
	Tr-CMC	0.36 (0.00)	0.39 (0.00)
	Tr-MC	0.22 (0.00)	0.35 (0.01)
	(baseline)	0.41 (0.00)	0.41 (0.00)
Movielens (100K)	DTr-CMC	0.39 (0.00)	0.38 (0.00)
	Fro-CMC	**0.41 (0.00)**	**0.41 (0.01)**
	Fro-MC	0.21 (0.01)	0.38 (0.01)
	Tr-CMC	0.40 (0.00)	0.40 (0.00)
	Tr-MC	0.12 (0.00)	0.38 (0.00)
	(baseline)	0.35 (0.00)	0.35 (0.00)

Results. The results are compiled in Table 3. In Task 1, by comparing the results between CMC methods and their corresponding MC methods, we conclude that the CMC methods have improved the ability to recover clipped values in real-world data as well. In Task 2, the CMC methods show better performance for predicting entries of the maximum value of rating than their MC counterparts.

Interestingly, we obtain the performance improvement by only changing the loss function to be robust to ceiling effects and without changing the model complexity (such as introducing an ordinal regression model). The computation time of the proposed methods are reported in Appendix C.

7 Conclusion

In this paper, we showed the first result of exact recovery guarantee for the novel problem of clipped matrix completion. We proposed practical algorithms as well as a theoretically-motivated regularization term. We showed that the clipped matrix completion methods obtained by modifying ordinary matrix completion methods are more robust to clipped data, through numerical experiments. A future work is to specialize our theoretical analysis to discrete data to analyze the ability of quantized matrix completion methods for recovering discrete data from ceiling effects.

Acknowledgments

We would like to thank Ikko Yamane, Han Bao, and Liyuan Xu, for helpful discussions. MS was supported by the International Research Center for Neurointelligence (WPI-IRCN) at The University of Tokyo Institutes for Advanced Study.

References

Austin, P. C., and Brunner, L. J. 2003. Type I error inflation in the presence of a ceiling effect. *The American Statistician* 57(2):97–104.

Austin, P. C.; Escobar, M.; and Kopec, J. A. 2000. The use of the Tobit model for analyzing measures of health status. *Quality of Life Research* 9(8):901–910.

Avron, H.; Kale, S.; Kasiviswanathan, S. P.; and Sindhwani, V. 2012. Efficient and practical stochastic subgradient descent for nuclear norm regularization. In *Proceedings of the 29th International Conference on Machine Learning*, 1231–1238.

Benjamin, R. 2005. A ceiling effect in traditional classroom foreign language instruction: Data from Russian. *The Modern Language Journal* 89(1):3–18.

Bhaskar, S. A. 2016. Probabilistic low-rank matrix completion from quantized measurements. *Journal of Machine Learning Research* 17(60):1–34.

Bhojanapalli, S., and Jain, P. 2014. Universal matrix completion. In *Proceedings of the 31st International Conference on Machine Learning*, 1881–1889.

Boucheron, S.; Lugosi, G.; and Massart, P. 2013. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford: Oxford University Press, 1st edition.

Cai, J.-F.; Cands, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4):1956–1982.

Cands, E. J., and Plan, Y. 2010. Matrix completion with noise. *Proceedings of the IEEE* 98(6):925–936.

Cands, E. J., and Recht, B. 2009. Exact matrix completion via convex optimization. *Foundations of Computational mathematics* 9(6):717–772.

Cands, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM* 58(3):1–37.

Catherine, L.; Nancy, Y.; Jasvir, M.; Marjorie, M.; and M., F. B. 2004. Revised versions of the Childhood Health Assessment Questionnaire (CHAQ) are more sensitive and suffer less from a ceiling effect. *Arthritis Care & Research* 51(6):881–889.

Chen, Y.; Bhojanapalli, S.; Sanghavi, S.; and Ward, R. 2015. Completing any low-rank matrix, provably. *Journal of Machine Learning Research* 16(1):2999–3034.

Cox, D. R., and Oakes, D. 1984. *Analysis of Survival Data*. New York: Chapman and Hall.

Davenport, M. A.; Plan, Y.; Van Den Berg, E.; and Wootters, M. 2014. 1-bit matrix completion. *Information and Inference: A Journal of the IMA* 3(3):189–223.

DeMaris, A. 2004. *Regression with Social Data: Modeling Continuous and Limited Response Variables*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley-Interscience.

Greene, W. H. 2012. *Econometric Analysis*. Boston: Prentice Hall, 7th edition.

Gross, D. 2011. Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory* 57(3):1548–1566.

Guo, G.; Zhang, J.; and Yorke-Smith, N. 2013. A novel Bayesian similarity measure for recommender systems. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, 2619–2625.

Horn, R. A. 1995. Norm bounds for Hadamard products and an arithmetic-geometric mean inequality for unitarily invariant norms. *Linear Algebra and its Applications* 223:355–361.

Jain, P.; Netrapalli, P.; and Sanghavi, S. 2013. Low-rank matrix completion using alternating minimization. In *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing*, 665–674.

Kaplan, C. 1992. Ceiling effects in assessing high-IQ children with the WPPSIR. *Journal of Clinical Child Psychology* 21(4):403–406.

Keshavan, R. H.; Montanari, A.; and Oh, S. 2010. Matrix completion from noisy entries. *Journal of Machine Learning Research* 11(Jul):2057–2078.

Kirly, F. J.; Theran, L.; and Tomioka, R. 2015. The algebraic combinatorial approach for low-rank matrix completion. *Journal of Machine Learning Research* 16(1):1391–1436.

Kohler, J. M., and Lucchi, A. 2017. Sub-sampled Cubic Regularization for Non-convex Optimization. In *Proceedings of the 34th International Conference on Machine Learning*, 1895–1904.

Lan, A. S.; Studer, C.; and Baraniuk, R. G. 2014. Matrix recovery from quantized and corrupted measurements. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 4973–4977.

Ledoux, M., and Talagrand, M. 1991. *Probability in Banach Spaces: Isoperimetry and Processes*. Berlin: Springer.

Lee, D. D., and Seung, H. S. 2001. Algorithms for Non-negative Matrix Factorization. In *Advances in Neural Information Processing Systems 13*. 556–562.

Liu, Z., and Vandenberghe, L. 2010. Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications* 31(3):1235–1256.

Liu, G.; Liu, Q.; and Yuan, X. 2017. A New Theory for Matrix Completion. In *Advances in Neural Information Processing Systems 30*. 785–794.

Mareek, J.; Richtrik, P.; and Tak, M. 2017. Matrix completion under interval uncertainty. *European Journal of Operational Research* 256(1):35–43.

Marecek, J.; Maroulis, S.; Kalogeraki, V.; and Gunopulos, D. 2018. Low-rank methods in event detection. *arXiv:1802.03649*.

Recht, B.; Fazel, M.; and Parrilo, P. 2010. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review* 52(3):471–501.

Recht, B. 2011. A simpler approach to matrix completion. *Journal of Machine Learning Research* 12(Dec):3413–3430.

Rodrigues, S. d. L. L.; Rodrigues, R. C. M.; Sao-Joao, T. M.; Pavan, R. B. B.; Padilha, K. M.; Gallani, M.-C.; Rodrigues, S. d. L. L.; Rodrigues, R. C. M.; Sao-Joao, T. M.; Pavan, R. B. B.; Padilha, K. M.; and Gallani, M.-C. 2013. Impact of the disease: Acceptability, ceiling and floor effects and reliability of an instrument on heart failure. *Revista da Escola de Enfermagem da USP* 47(5):1090–1097.

Salkind, N. J., ed. 2010. *Encyclopedia of Research Design*. Thousand Oaks, Calif: SAGE Publications.

Schnedler, W. 2005. Likelihood estimation for censored random vectors. *Econometric Reviews* 24(2):195–217.

Toh, K.-C., and Yun, S. 2010. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pacific Journal of Optimization* 6:615–640.

Tropp, J. A. 2012. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics* 12(4):389–434.

Voutilainen, A.; Pitkaho, T.; Kvist, T.; and Vehvilinen-Julkunen, K. 2016. How to ask about patient satisfaction? The visual analogue scale is less vulnerable to confounding factors and ceiling effect than a symmetric Likert scale. *Journal of Advanced Nursing* 72(4):946–957.

Yaginuma, H.; Kawai, S.; Tabata, K. V.; Tomiyama, K.; Kakizuka, A.; Komatsuzaki, T.; Noji, H.; and Imamura, H. 2014. Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Scientific Reports* 4:6522.