Smoothed Analysis of Moore-Penrose Inversion

Peter Bürgisser[*] and Felipe Cucker[†]

Abstract. We perform a smoothed analysis of the condition number of rectangular matrices. We prove that, asymptotically, the expected value of this condition number depends only of the elongation of the matrix, and not on the center and variance of the underlying probability distribution.

1 Introduction

The most widely used extension to rectangular matrices of the notion of inverse of square matrices is the so called Moore-Penrose inverse. For a full rank matrix $A \in \mathbb{R}^{m \times n}$ this is defined as $A^\dagger := (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}}$ if $m \geq n$, and as $A^\dagger := A^{\mathrm{T}} (A A^{\mathrm{T}})^{-1}$, otherwise. Immediate applications of $A^\dagger$ include the solution of least square problems

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|^2, \tag{1}$$

with $b \in \mathbb{R}^m$ and $m > n$, or of smallest solutions of underdetermined systems

$$\min_{x | Ax = b} \|x\|^2 \tag{2}$$

when $n > m$. In both cases, the solution is given by $x = A^\dagger b$. Well known results in error analysis show that the accuracy in the computation of $A^\dagger$, or in the computation of the solution x for the problems above, crucially depends on the *condition number* $\kappa(A) := \|A\| \|A^\dagger\|$ of A, where $\|A\|$ denotes the spectral norm (see [14, Ch. 19]). Accuracy analysis is not the only source of interest in $\kappa(A)$. Algorithms such as the conjugate gradient method produce approximate solutions of linear systems $Px = c$ —here $P \in \mathbb{R}^{m \times m}$ is a positive definite matrix and $c \in \mathbb{R}^m$— with a number of iterations proportional to $\sqrt{\kappa(P)}$ and, in many cases, the matrix P has been obtained as $P = A A^{\mathrm{T}}$ for some matrix $A \in \mathbb{R}^{m \times n}$. In those cases, $\sqrt{\kappa(P)} = \kappa(A)$ and one is again interested in the latter, this time by complexity considerations.

The condition number $\kappa(A)$ is not directly readable from A, and its computation seems to require that of $A^\dagger$. This is a common situation in numerical analysis. A way out of it, proposed as early as 1951 by von Neumann and Goldstine [17] and more recently pioneered by Demmel [6] and Smale [20], consists of randomizing the matrix A —say, by endowing $\mathbb{R}^{m \times n}$ with a multivariate standard Gaussian distribution $N(0, \mathrm{I})$— and considering its condition number as a derived random variable.

In Chen and Dongarra [4] the following tail estimates on $\kappa(A)$ were shown for $A \in \mathbb{R}^{m \times n}$ with $n \geq m$: for $x \geq n - m + 1$ we have

$$\frac{1}{\sqrt{2\pi}} \left(\frac{1}{5x}\right)^{n-m+1} \leq \operatorname*{Prob}_{A \sim N(0,\mathrm{I})} \left\{ \kappa(A) \geq \frac{x}{1-\lambda} \right\} \leq \frac{1}{\sqrt{2\pi}} \left(\frac{7}{x}\right)^{n-m+1}. \qquad (3)$$

Moreover, the expectation $\mathbb{E}(\kappa(A))$ can be bounded as a function of the *elongation* $\frac{m-1}{n}$ only, independently of n. (We remark that this is not true for Demmel's scaled condition number $\|A\|_F \|A^\dagger\|$, compare [9].) More precisely, for a sequence (m_n) of integers such that $\lim_{n \to \infty} m_n/n = \lambda \in (0, 1)$ and a sequence of standard Gaussian random matrices $A_n \in \mathbb{R}^{m_n \times n}$, we have in almost sure convergence

$$\kappa(A_n) \xrightarrow{\text{a.s.}} \frac{1 + \sqrt{\lambda}}{1 - \sqrt{\lambda}}. \qquad (4)$$

This follows from Geman [10] and Silverstein [19] (see Edelman [8] for more precise results).

The above results provide theoretical reasons of why least squares problems such as (1) or underdetermined systems such as (2) are solved to great accuracy or why the conjugate gradient method is so efficient in practice. In fact, it follows from (4) that the expected number of iterations of the conjugate gradient method on the random input $P = AA^{\mathrm{T}}$ remains bounded in terms of the *elongation* m/n as $n \to \infty$ and $A \in \mathbb{R}^{m \times n}$ is standard Gaussian. Our main result stated below implies that this phenomenon is still true for any matrix that is only slightly perturbed.

The choice of $N(0, \mathrm{I})$ as underlying data distribution is pervasive in the *average-case analysis* of condition numbers (and other quantities occurring in numerical analysis). It has the virtue of simplicity as a first approach to understanding which condition numbers one may expect. But it has been criticized due to the loose relationship of the Gaussian $N(0, \mathrm{I})$ to the measures that may be governing data drawing in practice. In particular, it has been observed that the use of Gaussians may be 'optimistic' in the sense that they may put more probability mass on the instances where the values of the function ψ under consideration are small. Such an optimism would produce yield an expectation $\mathbb{E}(\psi)$ smaller than the true one.

An alternate, more conservative, form of analysis has been proposed by Spielman and Teng under the name of *smoothed analysis*. It replaces the Gaussian measure $N(0, \mathrm{I})$ by the measures $N(\overline{A}, \sigma^2 \mathrm{I})$ where $\overline{A}$ is arbitrary. The idea is then to replace the unlikely 'average data' by a (usually small) perturbation of any possible

2

occurring data. The rationale for this form of analysis is offered in a number of papers [21, 18, 22, 23] and we won't repeat it here in full. We note, nonetheless, that the local nature of randomization in smoothed analysis, coupled with its worst-case dependence on the input data, removes from smoothed analysis the possible optimism we mentioned above for average-case analysis. In recent years, different aspects of algorithm behavior for a variety of problems have been analyzed this way. These include condition numbers of square matrices with real [27] or $\{-1, 1\}$ coefficients [24], complexity of interior-point methods [7], and machine learning [1]. The typical satisfying result is *polynomial smoothed complexity* (see [23, Def. 2]), consisting of a bound of the form

$$\sup_{\overline{A}} \; \mathbb{E}_{A \sim N(\overline{A}, \sigma^2 \mathrm{I})} \; \psi(A) \le c\sigma^{-k_1} \mathsf{size}(A)^{k_2} \tag{5}$$

where ψ is the function whose behavior we are analyzing and c, k_1, k_2 are positive constants.

In this paper we provide a smoothed analysis for Moore-Penrose inversion, extending (3) from the average-case analysis to smoothed analysis. To state the results we need to introduce some notations. We assume $1 \le m \le n$ throughout the paper. For a standard Gaussian $X \in \mathbb{R}^{m \times n}$ we put

$$Q(m, n) := \frac{1}{\sqrt{n}} \; \mathbb{E}(\|X\|). \tag{6}$$

(Lemma 2.4 shows that $Q(m, n) \le 6$.) We define for $\lambda \in (0, 1)$ the quantity

$$c(\lambda) \; := \; \sqrt{\frac{1 + \lambda}{2(1 - \lambda)}}. \tag{7}$$

Note that $c(\lambda)$ is monotonically increasing, $\lim_{\lambda \to 0} c(\lambda) = \frac{1}{\sqrt{2}}$ and $\lim_{\lambda \to 1} c(\lambda) = \infty$. Further, for $1 \le m \le n$ and $0 < \sigma \le 1$, we define the *elongation* $\lambda := \frac{m-1}{n}$ and introduce the quantity

$$\zeta_\sigma(m, n) \; := \; \left(Q(m, n) + \frac{1}{\sigma\sqrt{n}} \right) c(\lambda)^{\frac{1}{n-m+1}}. \tag{8}$$

Our main result is the following tail bound on the condition number of rectangular matrices under local Gaussian perturbations.

Theorem 1.1 *Suppose that $\overline{A} \in \mathbb{R}^{m \times n}$ satisfies $\|\overline{A}\| \le 1$ and let $0 < \sigma \le 1$. Put $\lambda := \frac{m-1}{n}$. Then, for $z \ge \zeta_\sigma(m, n)$, we have*

$$\Prob_{A \sim N(\overline{A}, \sigma^2 \mathrm{I})} \left\{ \kappa(A) \ge \frac{ez}{1 - \lambda} \right\} \; \le \; 2c(\lambda) \left[\left(Q(m, n) + \sqrt{2\ln(2z)} + \frac{1}{\sigma\sqrt{n}} \right) \frac{1}{z} \right]^{n-m+1}.$$

3

Remark 1.2 1. The decay in z in this tail bound is the same as in (3) up to the logarithmic factor $\sqrt{\ln z}$. We believe that the latter is an artefact of our proof that could be omitted. In fact, the exponent $n - m + 1$ is just the codimension of the set $\Sigma := \{A \in \mathbb{R}^{m \times n} \mid \mathrm{rk}A < m\}$ of rank deficient matrices, cf. [12]. Moreover, it is known [14] that $\|A^\dagger\| = 1/\mathrm{dist}(A, \Sigma)$ where the distance is measured in the Euclidean norm. From the interpretation of $\mathrm{Prob}\{\kappa(A) \geq t\}$ as the volume of a tube around Σ, as discussed in [2], one would therefore expect a decay of order $1/z^{n-m+1}$.

2. When $\sigma = 1$ and $\overline{A} = 0$, Theorem 1.1 yields tail bounds for the usual average case. One may therefore compare these bounds with (3). In doing so, we see that the bound in Theorem 1.1 has the additional factor $c(\lambda)$ (going to ∞ for $\lambda \to 1$). However, we note that the bound (3) holds only for $x = ez \geq n - m + 1$, while our bound holds for any $z \geq \zeta_\sigma(m, n)$. Furthermore, if we fix $\lambda \in (0, 1)$ and let (m_n) be a sequence of positive integers such that $\lim m_n/n = \lambda$, it follows from [10] that

$$\lim_{n \to \infty} Q(m_n, n) = 1 + \sqrt{\lambda}.$$

This implies that $\lim_{n \to \infty} \zeta_\sigma(m_n, n) = 1 + \sqrt{\lambda}$ for fixed $\sigma \in (0, 1]$ and, in particular, that $\zeta_\sigma(m_n, n) \leq 2$ for sufficiently large n . That is, for large n, the tail bound in Theorem 1.1 is valid for any $z \geq 2$.

Theorem 1.1 easily implies the following bound on expectations.

Corollary 1.3 *For all $\lambda_0 \in (0, 1)$ there exists n_0 such that for all $1 \leq m \leq n$ such that $\lambda = \frac{m-1}{n} \leq \lambda_0$ and $n \geq n_0$ we have for all σ with $\frac{1}{\sqrt{m}} \leq \sigma \leq 1$, and all $\overline{A} \in \mathbb{R}^{m \times n}$ with $\|\overline{A}\| \leq 1$, that*

$$\mathop{\mathbb{E}}_{A \sim N(\overline{A}, \sigma^2 \mathrm{I})} (\kappa(A)) \leq \frac{20.1}{1 - \lambda}.$$

As for the average-case analysis, this bound is independent of n and depends only on the bound λ_0 on the elongation. Thus we have a bound of type (5) with $k_2 = 0$. Surprisingly, the smoothed complexity bound in Corollary 1.3 is also independent of σ. We thus add reasons —and we will become more specific in Section 4— to the current understanding of the accuracy in least squares or underdetermined system solving or the complexity of the conjugate gradient method.

A first approach to the smoothed analysis of Moore-Penrose inversion appears in [5]. The bounds obtained in that paper are worse by an order of magnitude than those we obtain here. In Section 5 we compare these bounds with ours as well as with actual averages obtained, for specific values of n, m and σ, in numerical simulations.

Our proof techniques are an extension of methods employed by Sankar et al. [18].

2 Preliminaries

2.1 Some definitions and notation

The *spectral norm* of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as $\|A\| := \sup_{\|x\|=1} \|Ax\|$, where $\|x\|$ denotes the Euclidean norm. The *Frobenius norm* of A is defined as the Euclidean norm of A when interpreted as a vector.

Suppose that $A \in \mathbb{R}^{m \times n}$ is of maximal rank and $m \leq n$. The *Moore-Penrose inverse* of A is defined as $A^\dagger := A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1} \in \mathbb{R}^{n \times m}$. It can also be characterized as follows. For any $v \in \mathbb{R}^m$ the vector $w = A^\dagger v$ is orthogonal to the kernel of A and satisfies $Aw = v$. The *condition number* $\kappa(A)$ is defined as $\kappa(A) := \|A\| \cdot \|A^\dagger\|$.

Let $\overline{A} \in \mathbb{R}^{m \times n}$ and $\sigma > 0$. The *isotropic normal distribution* $N(\overline{A}, \sigma \mathrm{I})$ with center $\overline{A}$ and covariance matrix $\sigma^2 \mathrm{I}$ is the probability distribution on $\mathbb{R}^{m \times n}$ with the density

$$\rho_{\overline{A}, \sigma}(A) := \frac{1}{(2\pi)^{\frac{mn}{2}}} \, e^{-\frac{\|A-\overline{A}\|_F^2}{2\sigma^2}}.$$

Lemma 2.1 *For* $\lambda \in (0,1)$ *we have* $\lambda^{-\frac{\lambda}{1-\lambda}} \leq e$.

PROOF. Writing $u = 1/\lambda$ the assertion is equivalent to $u^{\frac{1}{u-1}} \leq e$ or $u \leq e^{u-1}$, which is certainly true for $u \geq 1$. $\square$

2.2 Concentration on spheres

Let $\mathbb{S}^{m-1} := \{x \in \mathbb{R}^m \mid \|x\| = 1\}$ denote the unit sphere in $\mathbb{R}^m$. We denote by $\mathcal{O}_{m-1}$ its volume, which is given by $\mathcal{O}_{m-1} = 2\pi^{m/2}/\Gamma(\frac{m}{2})$.

The following estimate tells us how likely a random point on $\mathbb{S}^{m-1}$ will lie in a fixed spherical cap.

Lemma 2.2 *Let* $u \in \mathbb{S}^{m-1}$ *be fixed,* $m \geq 2$. *Then, for all* $\xi \in [0,1]$,

$$\Prob_{v \sim U(\mathbb{S}^{m-1})} \{|u^{\mathrm{T}}v| \geq \xi\} \geq \sqrt{\frac{2}{\pi m}} \, (1-\xi^2)^{\frac{m-1}{2}}.$$

PROOF. We put $\theta = \arccos \xi$ and let $\mathsf{cap}(u,\theta)$ denote the spherical cap in $\mathbb{S}^{m-1}$ with center u and angular radius θ. Using the bounds in Lemmas 2.1 and 2.2 of [3] we get

$$\Prob_{v \sim U(\mathbb{S}^{m-1})} \{|u^{\mathrm{T}}v| \geq \xi\} = \frac{2 \, \mathsf{vol} \, \mathsf{cap}(u,\theta)}{\mathsf{vol} \, \mathbb{S}^{m-1}} \geq \frac{2\mathcal{O}_{m-2}}{\mathcal{O}_{m-1}} \frac{(1-\xi^2)^{\frac{m-1}{2}}}{(m-1)}.$$

Using the formula for $\mathcal{O}_{m-1}$ and the recursion $\Gamma(x+1) = x\Gamma(x)$ we have

$$\frac{\mathcal{O}_{m-2}}{\mathcal{O}_{m-1}} = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)} = \frac{1}{\sqrt{\pi}} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m-1}{2}\right)} \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m+1}{2}\right)} = \frac{m-1}{2\sqrt{\pi}} \frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m+1}{2}\right)}.$$

The assertion follows now from the estimate

$$\frac{\Gamma\left(\frac{m}{2}\right)}{\Gamma\left(\frac{m+1}{2}\right)} \geq \sqrt{\frac{2}{m}}. \tag{9}$$

This estimate can be quickly seen as follows. Suppose that $Z \in \mathbb{R}^m$ is standard normal distributed. Using polar coordinates and the variable transformation $u = \rho^2/2$ we get

$$\begin{aligned}
\mathbb{E}(\|Z\|) &= \frac{\mathcal{O}_{m-1}}{(2\pi)^{\frac{m}{2}}} \int_0^\infty \rho^m e^{-\frac{\rho^2}{2}} d\rho = \frac{\mathcal{O}_{m-1}}{(2\pi)^{\frac{m}{2}}} 2^{\frac{m-1}{2}} \int_0^\infty u^{\frac{m-1}{2}} e^{-u} du \\
&= \frac{\mathcal{O}_{m-1}}{(2\pi)^{\frac{m}{2}}} 2^{\frac{m-1}{2}} \Gamma\left(\frac{m+1}{2}\right) = \sqrt{2} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})},
\end{aligned} \tag{10}$$

where we used the definition of the Gamma function for the second last equality. To complete the proof of (9) we note that $\mathbb{E}(\|Z\|) \leq \sqrt{\mathbb{E}(\|Z\|^2)} = \sqrt{m}$. $\qquad\square$

For later use we note that (10) implies

$$\frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} = \frac{\Gamma\left(\frac{m+2}{2}\right)}{\Gamma\left(\frac{m}{2}\right)} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m+2}{2}\right)} = \frac{m}{2} \frac{\Gamma\left(\frac{m+1}{2}\right)}{\Gamma\left(\frac{m+2}{2}\right)} \geq \frac{m}{2} \sqrt{\frac{2}{m+1}},$$

using (9) for the right-hand inequality. Therefore

$$\mathbb{E}(\|Z\|) \geq \frac{m}{\sqrt{m+1}}. \tag{11}$$

2.3 Large deviations

We will use a powerful large deviation result. Let $F : \mathbb{R}^N \to \mathbb{R}$ be a Lipschitz continous function with Lipschitz constant L, so that $|F(x) - F(y)| \leq L\|x - y\|$ for all $x, y \in \mathbb{R}^N$, where $\| \ \|$ denotes the Euclidean norm. Now suppose that $x \in \mathbb{R}^N$ is a standard Gaussian random vector such that $\mathbb{E}(F(x))$ exists. Then it is known [16, (1.4)] that for all $t > 0$

$$\text{Prob}\{F(x) \geq \mathbb{E}(F) + t\} \leq e^{-\frac{t^2}{2L^2}}. \tag{12}$$

(We note that in [16, (1.4)] this is only stated for the median, but the inequality holds as well for the expectation. See also [15].)

2.4 A bound on the expected spectral norm

The function $\mathbb{R}^{m \times n} \to \mathbb{R}$ mapping a matrix X to its spectral norm $\|X\|$ is Lipschitz continuous with Lipschitz constant 1, as $\|X - Y\| \leq \|X - Y\|_F$. The concentration bound (12), together with (6), implies that for $t > 0$,

$$\text{Prob}\left\{\|X\| \geq Q(m,n)\sqrt{n} + t\right\} \leq e^{-\frac{t^2}{2}}. \tag{13}$$

This tail bound easily implies the following large deviation result.

6

Proposition 2.3 *Let* $\overline{A} \in \mathbb{R}^{m \times n}$ *with* $m \leq n$, $\|\overline{A}\| \leq 1$, *and* $\sigma \in (0, 1]$. *If* $A \in \mathbb{R}^{m \times n}$ *follows the law* $N(\overline{A}, \sigma^2 I)$, *then, for* $t > 0$,

$$\Prob_{A \sim N(\overline{A}, \sigma^2 I)} \left\{ \|A\| \geq Q(m, n)\sigma\sqrt{n} + t + 1 \right\} \leq e^{-\frac{t^2}{2\sigma^2}}.$$

PROOF.   We note that $\|A\| \geq Q(m, n)\sigma\sqrt{n} + t + 1$ implies that $\|A - \overline{A}\| \geq \|A\| - \|\overline{A}\| \geq Q(m, n)\sqrt{n} + t$. Moreover, if $A \in \mathbb{R}^{m \times n}$ follows the law $N(\overline{A}, \sigma^2 I)$, then $X := \frac{A - \overline{A}}{\sigma}$ is standard Gaussian in $\mathbb{R}^{m \times n}$. The assertion follows from (13). □

We derive now an upper bound on $Q(m, n)$. Such result should be well-known but we could not locate in the literature.

Lemma 2.4 *For* $n > 1$ *we have* $\sqrt{\frac{n}{n+1}} \leq Q(m, n) \leq 2\left(1 + \sqrt{\frac{2\ln(2m-1)}{n}} + \frac{1}{\sqrt{n}}\right) \leq 6.$

The proof relies on the following lemma.

Lemma 2.5 *Let* $r_1, \ldots, r_n$ *be independent random variables with nonnegative values such that* r_i^2 *is* χ^2-*distributed with* f_i *degrees of freedom. Then,*

$$\mathbb{E}\left(\max_{1 \leq i \leq n} r_i\right) \leq \max_{1 \leq i \leq n} \sqrt{f_i} + \sqrt{2\ln n} + 1.$$

PROOF.   We start by a large deviation estimate for χ^2-distributed random variables. Note that $\mathbb{R}^f \to \mathbb{R}$, $x \mapsto \|x\|$, is Lipschitz continuous with Lipschitz constant 1. From (12) we know that for standard Gaussian $x \in \mathbb{R}^n$ and all $t > 0$,

$$\Prob\{\|x\| \geq \mathbb{E}(\|x\|) + t\} \leq e^{-\frac{t^2}{2}}.$$

Since $\mathbb{E}(\|x\|) \leq \sqrt{\mathbb{E}(\|x\|^2)} = \sqrt{f}$, this implies for all $t > 0$,

$$\Prob\{\|x\| \geq \sqrt{f} + t\} \leq e^{-\frac{t^2}{2}}. \tag{14}$$

We suppose now that $r_1, \ldots, r_n$ are independent random variables with non-negative values such that r_i^2 is χ^2-distributed with f_i degrees of freedom. Put $f := \max_i f_i$. Equation (14) tells us that for all i and all $t > 0$,

$$\Prob\{r_i \geq \sqrt{f} + t\} \leq e^{-\frac{t^2}{2}}$$

and hence, by the union bound,

$$\Prob\left\{\max_{1 \leq i \leq n} r_i \geq \sqrt{f} + t\right\} \leq ne^{-\frac{t^2}{2}}.$$

7

For a fixed parameter $b \geq 1$ (to be determined later), this implies

$$
\begin{aligned}
\mathbb{E}(\max_{1 \leq i \leq n} r_i) &\leq \sqrt{f} + b + \int_{\sqrt{f}+b}^{\infty} \mathrm{Prob}\{\max_{1 \leq i \leq n} r_i \geq T\} \, dT \\
&= \sqrt{f} + b + \int_{b}^{\infty} \mathrm{Prob}\{\max_{1 \leq i \leq n} r_i \geq \sqrt{f} + t\} \, dt \\
&\leq \sqrt{f} + b + n \int_{b}^{\infty} e^{-\frac{t^2}{2}} \, dt.
\end{aligned}
$$

Using the well-known estimate

$$
\frac{1}{\sqrt{2\pi}} \int_{b}^{\infty} e^{-\frac{t^2}{2}} \, dt \leq \frac{1}{b\sqrt{2\pi}} e^{-\frac{b^2}{2}} \leq \frac{1}{\sqrt{2\pi}} e^{-\frac{b^2}{2}}
$$

we obtain

$$
\mathbb{E}(\max_{1 \leq i \leq n} r_i) \leq \sqrt{f} + b + n e^{-\frac{b^2}{2}}.
$$

Finally, choosing $b := \sqrt{2 \ln n}$ we get

$$
\mathbb{E}(\max_{1 \leq i \leq n} r_i) \leq \sqrt{f} + \sqrt{2 \ln n} + 1,
$$

as claimed. □

PROOF OF LEMMA 2.4. A general matrix $X \in \mathbb{R}^{m \times n}$ can be transformed into a bidiagonal matrix of the form

$$
Y := \begin{bmatrix}
v_n & & & & 0 & \cdots & 0 \\
w_{m-1} & v_{n-1} & & & \vdots & & \vdots \\
& \ddots & \ddots & & \vdots & & \vdots \\
& & w_1 & v_{n-m+1} & 0 & \cdots & 0
\end{bmatrix}
$$

with $v_i, w_j \geq 0$ by performing Householder transformations from the left and right hand side of X, cf. [11, §5.4.3]. In particular, $\|X\| = \|Y\|$. An analysis of this transformation shows that if we start with a standard Gaussian matrix X, then the $v_n, \ldots, v_{n-m+1}, w_{m-1}, \ldots, w_1$ are independent random variables such that v_i^2 and w_i^2 are χ^2-distributed with i degrees of freedom, cf. [19].

The spectral norm of Y is bounded by $\max_i v_i + \max_j w_j \leq 2r$, where r denotes the maximum of the values v_i and w_j. Lemma 2.5 implies that, for $n > 1$,

$$
\mathbb{E}(r) \leq \sqrt{n} + \sqrt{2 \ln(2m - 1)} + 1 \leq 3\sqrt{n}.
$$

This shows the claimed upper bound on $Q(m, n)$. For the lower bound we note that $\|Y\| \geq |v_n|$ which gives $\mathbb{E}(\|Y\|) \geq \mathbb{E}(|v_n|)$. The claimed lower bound now follows from (11), which states that $\mathbb{E}(|v_n|) \geq \sqrt{\frac{n}{n+1}}$. □

3 Proof of the main results

The main work consists of deriving tail bounds on $\|A^\dagger\|$, which is done in the next subsection.

3.1 Tail bounds for $\|A^\dagger\|$

Proposition 3.1 *Let $\overline{A} \in \mathbb{R}^{m \times n}$, $\sigma > 0$, and put $\lambda := \frac{m-1}{n}$. For random $A \sim N(\overline{A}, \sigma^2 I)$ we have, for any $t > 0$,*

$$\Prob_{A \sim N(\overline{A}, \sigma^2 I)} \left\{ \|A^\dagger\| \geq \frac{t}{1-\lambda} \right\} \leq c(\lambda) \left(\frac{e}{\sigma \sqrt{n}\, t} \right)^{(1-\lambda)n}.$$

We first show the following result.

Proposition 3.2 *For all $v \in \mathbb{S}^{m-1}$, $\overline{A} \in \mathbb{R}^{m \times n}$, $\sigma > 0$, and $\xi > 0$ we have*

$$\Prob_{A \sim N(\overline{A}, \sigma^2 I)} \left\{ \|A^\dagger v\| \geq \xi \right\} \leq \frac{1}{(\sqrt{2\pi})^{n-m+1}} \frac{\mathcal{O}_{n-m}}{n-m+1} \left(\frac{1}{\sigma \xi} \right)^{n-m+1}.$$

PROOF.   We first claim that, because of unitary invariance, we may assume that $v = e_m := (0, \ldots, 0, 1)$. To see this, take $\Phi \in U(m)$ such that $v = \Phi e_m$. Consider the isometric map $A \mapsto B = \Phi^{-1} A$ which transforms the density $\rho_{\overline{A}, \sigma}(A)$ into a density of the same form, namely $\rho_{\Phi^{-1}\overline{A}, \sigma}(B)$. Thus the assertion for e_m and random B implies the assertion for v and A, noting that $A^\dagger v = B^\dagger e_m$. This proves the claim.

We are going to characterize the norm of $w := A^\dagger e_m$ in a geometric way. Let a_i denote the ith row of A. Almost surely, the rows $a_1, \ldots, a_m$ are linearly independent; hence, we assume so in what follows. Let

$$R := \mathsf{span}\{a_1, \ldots, a_m\}, \quad S := \mathsf{span}\{a_1, \ldots, a_{m-1}\}.$$

Let $S^\perp$ denote the orthogonal complement of S in $\mathbb{R}^n$. We decompose $a_m = a_m^\perp + a_m^S$, where $a_m^\perp$ denotes the orthogonal projection of a_m onto $S^\perp$ and $a_m^S \in S$. Then $a_m^\perp \in R$ since both a_m and a_m^S are in R. It follows that $a_m^\perp \in R \cap S^\perp$.

We claim that $w \in R \cap S^\perp$ as well. Indeed, note that R equals the orthogonal complement of the kernel of A in $\mathbb{R}^n$. Therefore, by definition of the Moore-Penrose inverse, $w = A^\dagger e_m$ lies in R. Moreover, since $AA^\dagger = I$, we have $\langle w, a_i \rangle = 0$ for $i = 1, \ldots, m-1$ and hence $w \in S^\perp$ as well.

It is immediate to see that $\dim R \cap S^\perp = 1$. It then follows that $R \cap S^\perp = \mathbb{R}w = \mathbb{R}a_m^\perp$. Since $\langle w, a_m \rangle = 1$, we get $1 = \langle w, a_m \rangle = \langle w, a_m^\perp \rangle = \|w\| \, \|a_m^\perp\|$ and therefore

$$\|A^\dagger e_m\| = \frac{1}{\|a_m^\perp\|}. \tag{15}$$

Let $A_m \in \mathbb{R}^{(m-1) \times n}$ denote the matrix obtained from A by omitting a_m. The density $\rho_{\overline{A}, \sigma}$ factors as $\rho_{\overline{A}, \sigma}(A) = \rho_1(A_n) \rho_2(a_n)$ where ρ_1 and ρ_2 denote the density

functions of $N(\overline{A}_m, \sigma^2\mathrm{I})$ and $N(\bar{a}_m, \sigma^2\mathrm{I})$, respectively (the meaning of $\overline{A}_m$ and $\bar{a}_m$ being clear). Fubini's Theorem combined with (15) yield, for $\xi > 0$,

$$
\underset{N(\overline{A},\sigma^2\mathrm{I})}{\mathrm{Prob}} \left\{ \|A^\dagger e_m\| \geq \xi \right\} \;=\; \int_{\|A^\dagger e_m\| \geq \xi} \rho_{\overline{A},\sigma^2\mathrm{I}}(A)\, dA \tag{16}
$$

$$
= \int_{A_m \in \mathbb{R}^{(m-1)\times n}} \rho_1(A_m) \cdot \left(\int_{\|a_m^\perp\| \leq 1/\xi} \rho_2(a_m)\, da_m \right) dA_m.
$$

To complete the proof it is sufficient to show the bound

$$
\int_{\|a_m^\perp\| \leq \frac{1}{\xi}} \rho_2(a_m)\, da_m \leq \frac{1}{(\sqrt{2\pi})^{n-m+1}} \frac{\mathcal{O}_{n-m}}{n-m+1} \left(\frac{1}{\sigma\xi} \right)^{n-m+1} \tag{17}
$$

for fixed, linearly independent $a_1, \dots, a_{m-1}$ and $\xi > 0$.

To show (17) note that $a_m^\perp \sim N(\bar{a}_m^\perp, \sigma^2\mathrm{I})$ in $S^\perp \simeq \mathbb{R}^{n-m+1}$ where $\bar{a}_m^\perp$ is the orthogonal projection of $\bar{a}_m$ onto $S^\perp$. Let B_r denote the ball of radius r in $\mathbb{R}^p$ centered at the origin. It is easy to see that $\mathsf{vol}\, B_r = \mathcal{O}_{p-1} r^p / p$. For any $\bar{x} \in \mathbb{R}^p$ and any $\sigma > 0$ we have

$$
\underset{x \sim N(\bar{x}, \sigma^2\mathrm{I})}{\mathrm{Prob}} \left\{ \|x\| \leq \varepsilon \right\} \;\leq\; \underset{x \sim N(0, \sigma^2\mathrm{I})}{\mathrm{Prob}} \left\{ \|x\| \leq \varepsilon \right\} \;=\; \frac{1}{(\sigma\sqrt{2\pi})^p} \int_{\|x\| \leq \varepsilon} e^{-\frac{\|x\|^2}{2\sigma^2}}\, dx
$$

$$
\overset{x = \sigma z}{=\!=} \frac{1}{(\sqrt{2\pi})^p} \int_{\|z\| \leq \frac{\varepsilon}{\sigma}} e^{-\frac{\|z\|^2}{2}}\, dz
$$

$$
\leq \frac{1}{(\sqrt{2\pi})^p} \, \mathsf{vol}\, B_{\frac{\varepsilon}{\sigma}} \;=\; \frac{1}{(\sqrt{2\pi})^p} \left(\frac{\varepsilon}{\sigma} \right)^p \mathsf{vol}\, B_1
$$

$$
= \frac{1}{(\sqrt{2\pi})^p} \left(\frac{\varepsilon}{\sigma} \right)^p \frac{\mathcal{O}_{p-1}}{p}.
$$

Taking $\bar{x} = \bar{a}_m^\perp$, $\varepsilon = \frac{1}{\xi}$, and $p = n - m + 1$ the claim (17) follows. $\qquad\square$

PROOF OF PROPOSITION 3.1. The proof is based on an idea in [18]. For $A \in \mathbb{R}^{m \times n}$ there exists $u_A \in \mathbb{S}^{m-1}$ such that $\|A^\dagger\| = \|A^\dagger u_A\|$. Moreover, for almost all A, the vector u_A is uniquely determined up to sign. Using the singular value decomposition it is easy to show that, for all $v \in \mathbb{S}^{m-1}$,

$$
\|A^\dagger v\| \geq \|A^\dagger\| \cdot |u_A^\mathrm{T} v|. \tag{18}
$$

Now take $A \sim N(\overline{A}, \sigma^2\mathrm{I})$ and $v \sim U(\mathbb{S}^{m-1})$ independently. Then, for any $s \in (0, 1)$ and $t > 0$ we have

$$
\underset{A,v}{\mathrm{Prob}} \left\{ \|A^\dagger v\| \geq t\sqrt{1 - s^2} \right\} \;\geq\; \underset{A,v}{\mathrm{Prob}} \left\{ \|A^\dagger\| \geq t \ \& \ |u_A^\mathrm{T} v| \geq \sqrt{1 - s^2} \right\}
$$

$$
= \underset{A}{\mathrm{Prob}} \left\{ \|A^\dagger\| \geq t \right\} \cdot \underset{A,v}{\mathrm{Prob}} \left\{ |u_A^\mathrm{T} v| \geq \sqrt{1 - s^2} \ \middle| \ \|A^\dagger\| \geq t \right\}
$$

$$
\geq \underset{A}{\mathrm{Prob}} \left\{ \|A^\dagger\| \geq t \right\} \cdot \sqrt{\frac{2}{\pi m}} \, s^{m-1},
$$

10

the last line by Lemma 2.2 with $\xi = \sqrt{1-s^2}$. Now we use Proposition 3.2 with $\xi = t\sqrt{1-s^2}$ to deduce that

$$
\begin{aligned}
\operatorname*{Prob}_{A}\left\{\|A^\dagger\| \geq t\right\} &\leq \sqrt{\frac{\pi m}{2}}\,\frac{1}{s^{m-1}}\operatorname*{Prob}_{A,v}\{\|A^\dagger v\| \geq t\sqrt{1-s^2}\} &(19)\\
&\leq \frac{\sqrt{m}}{2s^{m-1}}\frac{1}{(\sqrt{2\pi})^{n-m}}\frac{\mathcal{O}_{n-m}}{n-m+1}\left(\frac{1}{\sigma t\sqrt{1-s^2}}\right)^{n-m+1}.
\end{aligned}
$$

We next choose $s \in (0,1)$ to minimize the bound above. To do so amounts to maximize $(1-x)^{\frac{n-m+1}{2}}x^{\frac{m-1}{2}}$ where $x = s^2 \in (0,1)$, or yet, to maximize

$$
g(x) = \left((1-x)^{\frac{n-m+1}{2}}x^{\frac{m-1}{2}}\right)^{\frac{2}{n}} = (1-x)^{\frac{n-m+1}{n}}x^{\frac{m-1}{n}} = (1-x)^{1-\lambda}x^\lambda.
$$

We have $\frac{d}{dx}\ln g(x) = \frac{\lambda}{x} - \frac{1-\lambda}{1-x}$ with the only zero attained at $x^* = \lambda$.

Replacing s^2 by λ in (19) we obtain the bound

$$
\operatorname*{Prob}_{A}\left\{\|A^\dagger\| \geq t\right\} \leq \frac{\sqrt{\lambda n+1}}{2\lambda^{\frac{\lambda n}{2}}}\frac{1}{(\sqrt{2\pi})^{n-m}}\frac{\mathcal{O}_{n-m}}{(1-\lambda)n}\left(\frac{1}{\sigma t\sqrt{1-\lambda}}\right)^{(1-\lambda)n}.
$$

Lemma 2.1 implies

$$
\lambda^{-\frac{\lambda n}{2}} = \left(\lambda^{-\frac{\lambda}{2(1-\lambda)}}\right)^{(1-\lambda)n} \leq e^{\frac{(1-\lambda)n}{2}}.
$$

So we get

$$
\begin{aligned}
\operatorname*{Prob}_{A}\left\{\|A^\dagger\| \geq t\right\} &\leq \frac{\sqrt{\lambda n+1}}{2}\frac{1}{(\sqrt{2\pi})^{n-m}}\frac{\mathcal{O}_{n-m}}{(1-\lambda)n}\left(\frac{\sqrt{e}}{\sigma t\sqrt{1-\lambda}}\right)^{(1-\lambda)n}\\
&= \frac{\sqrt{\lambda n+1}}{2}\left(\frac{e}{1-\lambda}\right)^{\frac{(1-\lambda)n}{2}}\frac{1}{(\sqrt{2\pi})^{n-m}}\frac{\mathcal{O}_{n-m}}{(1-\lambda)n}\left(\frac{1}{\sigma t}\right)^{(1-\lambda)n}\\
&= \frac{1}{2(1-\lambda)}\sqrt{\lambda+\frac{1}{n}}\,\frac{1}{\sqrt{n}}\left(\frac{e}{1-\lambda}\right)^{\frac{(1-\lambda)n}{2}}\frac{\mathcal{O}_{n-m}}{(\sqrt{2\pi})^{n-m}}\left(\frac{1}{\sigma t}\right)^{(1-\lambda)n}\\
&\leq \frac{\sqrt{\lambda+1}}{2(1-\lambda)}\frac{1}{\sqrt{n}}\left(\frac{e}{1-\lambda}\right)^{\frac{(1-\lambda)n}{2}}\frac{2\pi^{\frac{n-m+1}{2}}}{\Gamma\left(\frac{n-m+1}{2}\right)(\sqrt{2\pi})^{n-m}}\left(\frac{1}{\sigma t}\right)^{(1-\lambda)n}\\
&= \frac{\sqrt{1+\lambda}}{1-\lambda}\frac{1}{\sqrt{n}}\left(\frac{e}{1-\lambda}\right)^{\frac{(1-\lambda)n}{2}}\frac{\sqrt{2\pi}}{\Gamma\left(\frac{n(1-\lambda)}{2}\right)2^{\frac{(1-\lambda)n}{2}}}\left(\frac{1}{\sigma t}\right)^{(1-\lambda)n}.
\end{aligned}
$$

We next estimate $\Gamma\left(\frac{(1-\lambda)n}{2}\right)$. To do so, recall Stirling's bound

$$
\sqrt{2\pi}x^{x+\frac{1}{2}}e^{-x} < \Gamma(x+1) < \sqrt{2\pi}x^{x+\frac{1}{2}}e^{-x+\frac{1}{12x}} \qquad \text{for all } x > 0
$$

11

which yields, using $\Gamma(x+1) = x\Gamma(x)$, the bound $\Gamma(x) > \sqrt{2\pi/x}\,(x/e)^x$. We use this with $x = \frac{(1-\lambda)n}{2}$ to obtain

$$\Gamma\left(\frac{(1-\lambda)n}{2}\right) \geq \sqrt{\frac{4\pi}{(1-\lambda)n}}\left(\frac{(1-\lambda)n}{2e}\right)^{\frac{(1-\lambda)n}{2}}.$$

Plugging this into the above we obtain (observe the crucial cancellation of $\sqrt{n}$)

$$\operatorname*{Prob}_A\left\{\|A^\dagger\| \geq t\right\}$$

$$\leq \sqrt{\frac{1+\lambda}{(1-\lambda)^2}}\,\frac{1}{\sqrt{n}}\left(\frac{e}{1-\lambda}\right)^{\frac{(1-\lambda)n}{2}}\sqrt{2\pi}\sqrt{\frac{(1-\lambda)n}{4\pi}}\left(\frac{e}{(1-\lambda)n}\right)^{\frac{(1-\lambda)n}{2}}\left(\frac{1}{\sigma t}\right)^{(1-\lambda)n}$$

$$= c(\lambda)\left(\frac{e}{1-\lambda}\right)^{(1-\lambda)n}\left(\frac{1}{n}\right)^{\frac{(1-\lambda)n}{2}}\left(\frac{1}{\sigma t}\right)^{(1-\lambda)n} = c(\lambda)\left(\frac{e}{\sigma\sqrt{n}(1-\lambda)t}\right)^{(1-\lambda)n},$$

which completes the proof of the proposition. $\qquad\square$

3.2 Proof of Theorem 1.1

To simplify notation we write $c := c(\lambda)$ and $Q := Q(m,n)$. Proposition 3.1 implies that for any $\varepsilon > 0$ we have

$$\operatorname*{Prob}_{A\sim N(\overline{A},\sigma^2 I)}\left\{\|A^\dagger\| \geq \frac{e}{1-\lambda}\frac{1}{\sigma\sqrt{n}}\left(\frac{c}{\varepsilon}\right)^{\frac{1}{(1-\lambda)n}}\right\} \leq \varepsilon. \tag{20}$$

Similarly, letting $\varepsilon = e^{-\frac{t^2}{2\sigma^2}}$ in Proposition 2.3 and solving for t we deduce that, for any $\varepsilon \in (0,1]$,

$$\operatorname{Prob}\left\{\|A\| \geq Q\sigma\sqrt{n} + \sigma\sqrt{2\ln\frac{1}{\varepsilon}} + 1\right\} \leq \varepsilon. \tag{21}$$

We conclude that

$$\operatorname*{Prob}_{A\sim N(\overline{A},\sigma^2 I)}\left\{\kappa(A) \geq \frac{ez(\varepsilon)}{1-\lambda}\right\} \leq 2\varepsilon, \tag{22}$$

where we have have set, for $\varepsilon \in (0,1]$,

$$z(\varepsilon) := \left(Q + \sqrt{\frac{2}{n}\ln\frac{1}{\varepsilon}} + \frac{1}{\sigma\sqrt{n}}\right)\left(\frac{c}{\varepsilon}\right)^{\frac{1}{(1-\lambda)n}}. \tag{23}$$

We note that $z(1) = \zeta := \zeta_\sigma(m,n)$, cf. Equation (8). Moreover, $\lim_{\varepsilon\to 0} z(\varepsilon) = \infty$ and z is decreasing in the interval $(0,1]$. Hence, for $z \geq \zeta$, there exists $\varepsilon = \varepsilon(z) \in (0,1]$ such that $z = z(\varepsilon)$.

12

We need to upper bound $\varepsilon(z)$ as a function of z. To do so, we start with a weak lower bound on $\varepsilon(z)$ and claim that

$$\frac{1}{n} \ln \frac{1}{\varepsilon} \leq \ln(2z(\varepsilon)). \tag{24}$$

To show this, recall that $Q \geq \sqrt{\frac{n}{n+1}} \geq \frac{1}{\sqrt{2}}$ due to Lemma 2.4. Hence $\zeta \geq Q \geq 1/\sqrt{2}$ and it follows that $\sqrt{2}z \leq 1$ for $z \geq \zeta$. Thus, Equation (23) implies that

$$z(\varepsilon) \geq \frac{1}{\sqrt{2}} \left(\frac{c}{\varepsilon}\right)^{\frac{1}{(1-\lambda)n}}.$$

Using $c \geq \frac{1}{\sqrt{2}}$ we get

$$(\sqrt{2}z)^n \geq (\sqrt{2}z)^{(1-\lambda)n} \geq \frac{c}{\varepsilon} \geq \frac{1}{\sqrt{2}\,\varepsilon}.$$

Hence $(2z)^n \geq 1/\varepsilon$, which shows the claimed inequality (24).

Using the bound (24) in Equation (23) we get, again writing $z = z(\varepsilon)$, that

$$z \leq \left(Q + \sqrt{2\ln(2z)} + \frac{1}{\sigma\sqrt{n}}\right) \left(\frac{c}{\varepsilon}\right)^{\frac{1}{(1-\lambda)n}},$$

which means

$$\varepsilon \leq c\left[\left(Q + \sqrt{2\ln(2z)} + \frac{1}{\sigma\sqrt{n}}\right)\frac{1}{z}\right]^{(1-\lambda)n}.$$

By (22) this completes the proof. $\qquad\square$

3.3 Proof of Corollary 1.3

Fix $\lambda_0 \in (0, 1)$ and put $c := c(\lambda_0)$. Suppose that $m \leq n$ satisfy $\lambda = (m-1)/n \leq \lambda_0$. Then $n-m+1 = (1-\lambda)n \geq (1-\lambda_0)n$ and in order to have $n-m$ sufficiently large it suffices to require that n is sufficiently large. Thus, $c^{\frac{1}{n-m+1}} \leq 1.1$ if n is sufficiently large. Similarly, because of Lemma 2.4, $Q(m,n) \leq 2.1$ for large enough n. This implies that, for $\frac{1}{\sqrt{m}} \leq \sigma \leq 1$, we have

$$Q(m,n) + \frac{1}{\sigma\sqrt{n}} \leq 2.1 + \frac{1}{\sigma\sqrt{n}} \leq 2.1 + \sqrt{\frac{m}{n}} \leq 2.1 + \sqrt{\lambda_0 + \frac{1}{n}} \leq 3.1,$$

provided n is large enough. Then $\zeta_\sigma(m,n) \leq 3.1 \cdot 1.1 = 3.41$.

By Theorem 1.1, the random variable $Z := (1-\lambda)\kappa(A)/e$ satisfies, for any $\overline{A}$ with $\|\overline{A}\| \leq 1$ and any $z \geq 3.41$,

$$\Prob_{A\sim N(\overline{A},\sigma^2 I)} \{Z \geq z\} \leq 2c\left[\left(Q(m,n) + \sqrt{2\ln(2z)} + \frac{1}{\sigma\sqrt{n}}\right)\frac{1}{z}\right]^{n-m+1}$$

$$\leq 2c\left[\left(3.1 + \sqrt{2\ln(2z)}\right)\frac{1}{z}\right]^{n-m+1}.$$

13

Since $3.1 + \sqrt{2\ln(2z)} \le e\sqrt{z}$ for $z \ge 4$ we deduce that, for all such z,

$$\operatorname*{Prob}_{A \sim N(\overline{A}, \sigma^2 \mathrm{I})} \{Z \ge z\} \ \le \ 2c \left(\frac{e}{\sqrt{z}}\right)^{n-m+1}.$$

Using this tail bound to compute $\mathbb{E}(Z)$ we get

$$
\mathbb{E}(Z) \quad = \quad \int_0^\infty \operatorname{Prob}\{Z \ge z\}\, dz \ \le \ e^2 + 2c \int_{e^2}^\infty \left(\frac{e^2}{z}\right)^{\frac{n-m+1}{2}} dz
$$

$$
\stackrel{z = e^2 y}{=} \quad e^2 + 2c \int_1^\infty \left(\frac{1}{y}\right)^{\frac{n-m+1}{2}} e^2 dy \ = \ e^2 + \frac{4ce^2}{n-m-1}.
$$

We can now conclude since

$$\mathbb{E}((1-\lambda)\kappa(A)) = \mathbb{E}(eZ) = e\,\mathbb{E}(Z) \le e^3 + \frac{4ce^3}{n-m-1} \le 20.1$$

the inequality, again, by taking n large enough. $\qquad\qquad\square$

4 Applications

We next briefly discuss the two applications of our main result mentioned in the introduction.

4.1 Accuracy of Linear Least Squares

Recall the problem (1) described in the introduction, namely, to compute the minimum of $\|Ax - b\|^2$ over $x \in \mathbb{R}^n$ for given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ (with $m > n$). It is well known that the loss of precision $\mathsf{LoP}(A^\dagger b)$ —that is, the number of correct digits in the entries of the data (A, b) minus the same number for the computed solution $A^\dagger b$— satisfies (cf. [26] and [14, Ch. 19])

$$\mathsf{LoP}(A^\dagger b) \le \log mn^{3/2} + 2\log \kappa(A) + \mathcal{O}(1).$$

Corollary 1.3, combined with Jensen's inequality, implies that $\mathbb{E}(\log \kappa(A)) \le \log(20.1/(1-\lambda)) = \mathcal{O}(1)$ under the assumptions stated in the corollary. Hence for sufficiently elongated, large matrices $\overline{A}$, the expected loss of precision in the computation of the solution $A^\dagger b$ over all small perturbations A of $\overline{A}$ is dominated by the term $\log mn^{3/2}$.

4.2 Complexity of the Conjugate Gradient Method

If $P \in \mathbb{R}^{m \times m}$ is a symmetric positive definite matrix and $c \in \mathbb{R}^m$, the system $Px = c$ can be solved by the Conjugate Gradient Method (CGM), cf. [13]. This is an iterative algorithm which performs at most m iterations but may require less.

Indeed, it is known (see, e.g., [25, Lecture 38]) that an ε-approximation of the solution x can be computed in at most $\frac{1}{2}\sqrt{\kappa(P)}|\ln\varepsilon|$ iterations (ε measures the relative error of the approximation with respect to the Euclidean norm).

In many cases the matrix P arises as $P = AA^{\mathrm{T}}$ for some matrix $A \in \mathbb{R}^{m\times n}$ with $n > m$. If A is standard Gaussian distributed, then the resulting distribution of P, called *Wishart distribution*, has been extensively studied in multivariate statistics. However, in our case of interest, A is noncentered and much less is known about the resulting distribution of P (called noncentral Wishart). Fortunately, using the fact that $\sqrt{\kappa(P)} = \kappa(A)$, we can directly apply our tail bounds for $\kappa(A)$ for a noncentral, isotropic Gaussian distribution of A, to derive bounds for the expected number of iterations of CGM.

To do so we use again Corollary 1.3. It shows that for all $\lambda_0 \in (0,1)$ and all $0 < \sigma \leq 1$ there exists n_0 such that for all $1 \leq m < n$ we have

$$\sup_{\|\overline{A}\|\leq 1} \mathop{\mathbb{E}}_{A\sim N(\overline{A},\sigma^2 \mathrm{I})} (\kappa(A)) \leq \frac{20.1}{1-\lambda},$$

provided $\lambda = \frac{m-1}{n} \leq \lambda_0$ and $n \geq n_0$. It follows that if P is obtained as AA^{T} for a large, elongated, rectangular matrix A then, we should expect to compute a solution with the desired accuracy with about $\frac{1}{2}\frac{20.1}{1-\lambda}|\ln\varepsilon|$ iterations. It is known that each iteration of CGM takes $6n^2 + \mathcal{O}(n)$ arithmetic operations. Therefore, the expected cost of running CGM on P is

$$3n^2\frac{20.1}{1-\lambda}|\ln\varepsilon| + \mathcal{O}(n) = \frac{60.3n^2}{1-\lambda}|\ln\varepsilon| + \mathcal{O}(n).$$

The leading term in this expression is smaller than the $\frac{2}{3}n^3$ operations performed by Gaussian elimination as long as

$$\varepsilon \geq e^{-\frac{n(1-\lambda)}{91}}.$$

For large n (and λ not too close to 1) this bound produces very small values of ε and therefore, CGM yields, on the average (both for a Wishart distribution of data P and for Wishart perturbations of arbitrary data), remarkably good approximations of the solution $x = P^{-1}c$.

5   Some Numerical Simulations

Section 6 in [5] describes the result of numerical computations producing experimental values for $\mathbb{E}(\ln\kappa(A))$ (for certain choices of $\overline{A}$ and σ), which are denoted by $\mathbf{Avr}(\ln\kappa(A))$ and compared with the upper bound for $\mathbb{E}(\ln\kappa(A))$

$$\mu(m,n,\sigma) := \ln\left(m + \sigma m\sqrt{5n}\right) + \ln\frac{2.35}{\sigma} + \frac{1}{r} + \sqrt{\frac{e\pi}{5}}$$

15

obtained there. The data in Tables 1 to 4 is taken from [5]. Each row in these tables corresponds to a pair (m, n). For each row, 500 random matrices $A \in \mathbb{R}^{m \times n}$ were computed following the distribution $N(\sqrt{m}\,\overline{A}, \mathrm{I})$, where $\overline{A}$ was chosen as

$$\overline{A} := \frac{\mathsf{ones}(\mathsf{m}, \mathsf{n})}{\|\mathsf{ones}(\mathsf{m}, \mathsf{n})\|},$$

and $\mathsf{ones(m,n)}$ denotes the $m \times n$ matrix all of whose entries are 1. The column with header $\mathbf{Avr}\,(\ln \kappa(A))$ shows the empirical average of $\ln \kappa(A)$ for the 500 chosen random matrices A. Since $\kappa(A)$ is scale invariant we note that this corresponds to random matrices chosen from $N(\overline{A}, \sigma\mathrm{I})$, where $\sigma = 1/\sqrt{m}$.

m	n	$\mathbf{Avr}\,(\ln \kappa(A))$	$\mu(m, n, \sigma)$	$\ln(20.1/(1 - \lambda))$
10	15	1.88278226808667	7.73190477060415	
20	30	2.04718612539162	8.74083698937094	
40	60	2.13539482051851	9.75820027818245	4.0993321
80	120	2.19377719811291	10.78180469776403	
160	240	2.23119383890675	11.80997066079053	

Table 1: $n = 1.5m$.

m	n	$\mathbf{Avr}\,(\ln \kappa(A))$	$\mu(m, n, \sigma)$	$\ln(20.1/(1 - \lambda))$
5	10	1.28204418194521	6.35902343647518	
10	20	1.48669849397793	7.36178009761038	
20	40	1.59394635398509	8.37451330180407	
40	80	1.64896402420115	9.39470162365532	3.693866
80	160	1.69565973841311	10.42037692088400	
160	320	1.72154032592663	11.45004561375610	

Table 2: $n = 2m$.

m	n	$\mathbf{Avr}\,(\ln \kappa(A))$	$\mu(m, n, \sigma)$	$\ln(20.1/(1 - \lambda))$
10	25	1.24167342192086	7.46370799208199	
20	50	1.34213347902230	8.47908853717777	
40	100	1.40120155287858	9.50123344342563	3.511545
80	200	1.44120596017225	10.52833707967242	
160	400	1.45928497502137	11.55903912197539	

Table 3: $n = 2.5m$.

m	n	$\mathbf{Avr}\,(\ln \kappa(A))$	$\mu(m,n,\sigma)$	$\ln(20.1/(1-\lambda))$
5	15	0.98741849882614	6.37209092337754	
10	30	1.10550395287499	7.38102314214432	
20	60	1.18790345922560	8.39838643095583	
40	120	1.23914387557043	9.42199085053742	3.406185
80	240	1.27096561714092	10.45015681356392	
160	480	1.28600775609989	12.14829242876138	

Table 4: $n = 3m$.

In [5] it is observed that "one sees that when one fixes m and lets n increase the quantity $\mathbf{Avr}(\ln \kappa(A))$ decreases. This is in contrast with the behaviour of $\mu(m,n,\sigma)$. It appears that our methods are not sharp enough to capture the behaviour of $\mathbb{E}(\ln \kappa(A))$." Compare now with the results of the present paper. It follows from Corollary 1.3, by Jensen's inequality, that, for sufficiently large n, $\mathbb{E}(\ln \kappa(A)) \leq \ln \frac{20.1}{1-\lambda}$. But if m is held fixed then, when n increases, λ decreases and so does $\ln \frac{20.1}{1-\lambda}$.

One still observes a difference between the bound $\ln \frac{20.1}{1-\lambda}$ and the values of $\mathbf{Avr}\,(\ln \kappa(A))$. Part of this difference comes from the asymptotic character of this bound and the fact that our data is limited to $m \leq 160$. One sees on the tables that larger values of m would approach $\mathbf{Avr}\,(\ln \kappa(A))$ to $\ln \frac{20.1}{1-\lambda}$. We conjecture that, in addition to the possible loss of sharpness coming from the use of Jensen's inequality, the difference above is due to the roughness of the constant 20.1.

References

[1] D. Arthur and S. Vassilvitskii. Worst-case and smoothed analysis of the ICP algorithm, with an application to the k-means method. *SIAM J. Comput.*, 39(2):766–782, 2009.

[2] P. Bürgisser, F. Cucker, and M. Lotz. Smoothed analysis of complex conic condition numbers. *J. Math. Pures Appl. (9)*, 86(4):293–309, 2006.

[3] P. Bürgisser, F. Cucker, and M. Lotz. The probability that a small perturbation of a numerical analysis problem is difficult. *Mathematics of Computation*, 77:1559–1583, 2008.

[4] Z.-Z. Cheng and J.J. Dongarra. Condition numbers of Gaussian random matrices. *SIAM J. Matrix Anal. Appl.*, 27:603–620, 2005.

[5] F. Cucker, H. Diao, and Y. Wei. Smoothed analysis of some condition numbers. *Numer. Lin. Alg. Appl.*, 13:71–84, 2005.

[6] J. Demmel. The probability that a numerical analysis problem is difficult. *Math. Comp.*, 50:449–480, 1988.

[7] J. Dunagan, D.A. Spielman, and S.-H. Teng. Smoothed analysis of condition numbers and complexity implications for linear programming. To appear at *Math. Programming*. Preprint available at `http://arxiv.org/abs/cs/0302011v2`, 2009.

[8] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[9] A. Edelman. On the distribution of a scaled condition number. *Math. Comp.*, 58(197):185–190, 1992.

[10] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.

[11] G.H. Golub and C.F. Van Loan. *Matrix computations*, volume 3 of *Johns Hopkins Series in the Mathematical Sciences*. Johns Hopkins University Press, Baltimore, MD, 1983.

[12] J. Harris. *Algebraic Geometry*, volume 133 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1995.

[13] M.-R. Hestenes and E. Stiefel. Methods of Conjugate Gradients for Solving Linear Systems. *J. Res. Natl. Bur. Stand.* 49:409-436, 1952.

[14] N. J. Higham. *Accuracy and stability of numerical algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1996.

[15] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

[16] M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991.

[17] J. von Neumann and H.H. Goldstine. Numerical inverting matrices of high order, II. *Proc. Amer. Math. Soc.*, 2:188–202, 1951.

[18] A. Sankar, D.A. Spielman, and S.H. Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476 (electronic), 2006.

[19] J.W. Silverstein. The smallest eigenvalue of a large-dimensional Wishart matrix. *Ann. Probab.*, 13(4):1364–1368, 1985.

[20] S. Smale. Complexity theory and numerical analysis. In A. Iserles, editor, *Acta Numerica*, pages 523–551. Cambridge University Press, 1997.

[21] D.A. Spielman and S.-H. Teng. Smoothed analysis of algorithms: why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463 (electronic), 2004.

[22] D.A. Spielman and S.-H. Teng. Smoothed analysis of algorithms and heuristics. In *Foundations of Computational Mathematics, Santander 2005*, volume 331 of *Lecture Notes of the London Mathematical Society*, pages 274–342, 2006.

[23] D.A. Spielman and S.-H. Teng. Smoothed analysis: An attempt to explain the behavior of algorithms in practice. *Communications of the ACM*, 52(10):77–84, 2009.

[24] T. Tao and V. Vu. The condition number of a randomly perturbed matrix. In *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 248–255. ACM, New York, 2007.

[25] L.N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM, Philadelphia, PA, 1997.

[26] P.-A. Wedin. Perturbation theory for pseudo-inverses. *BIT*, 13: 217–232, 1973.

[27] M. Wschebor. Smoothed analysis of $\kappa(A)$. *J. Complexity*, 20(1):97–107, 2004.