DELFT UNIVERSITY OF TECHNOLOGY

REPORT 10-09

On the statistical properties of solutions of completely random linear systems.

Peter Sonneveld

ISSN 1389-6520

Reports of the Department of Applied Mathematical Analysis

Delft 2010

On the statistical properties of solutions of completely random linear systems.

Peter Sonneveld*

June 1, 2010

Abstract

In this report the probability distributions of $\boldsymbol{x}$ and $\|\boldsymbol{x}\|$ are derived, where $\boldsymbol{x}$ is the solution of a finite square, completely random linear system, i.e. a system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, of which all entries are stochastically independent, and standard Gaussian distributed stochastic variables.

The question on the statistical behaviour of $\|\boldsymbol{x}\|$ came up in [5], the convergence analysis of IDR(s), which is a recently developed, short-recurrence Krylov subspace based iterative solution method for large sparse non-symmetric linear systems of equations, ([6], [4]).

Keywords: Random matrices, Wishart distribution, IDR(s)

1 Introduction

In [6] IDR(s), a new short-recurrence Krylov subspace solution method for large sparse linear systems is introduced. Experiments with this method showed a convergence behaviour which was rather similar to the convergence of GMRES, a very stable solution method for the same class of linear systems. GMRES is a *long-recurrence* method, which means that the computing time as well as the memory load contain a component that is growing linearly with the iteration count. For large scale problems requiring many iteration steps for convergence, the GMRES procedure often turns out to be too expensive.

In general, a Krylov subspace method produces approximations $\boldsymbol{x}_n$ to the solution $\boldsymbol{x}$ of a linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, for which the residuals $\boldsymbol{r}_n = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_n$ are in the Krylov subspaces $\mathcal{K}^n(\boldsymbol{A}, \boldsymbol{r}_0) = \mathrm{span}(\boldsymbol{r}_0, \boldsymbol{A}\boldsymbol{r}_0, \ldots, \boldsymbol{A}^n\boldsymbol{r}_0)$, where $\boldsymbol{r}_0 = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}_0$, and $\boldsymbol{x}_0$ is an initial guess.

In [5] IDR(s) was compared to GMRES, and it turned out that for large values of the parameter s, the IDR(s)-residuals $\boldsymbol{r}_n$ and the GMRES-residuals $\widehat{\boldsymbol{r}}_n$ are related by

$$\boldsymbol{r}_n = \Omega_j(\boldsymbol{A})\,(\,\widehat{\boldsymbol{r}}_{n-j} + \|\widehat{\boldsymbol{r}}_{n-j}\|\boldsymbol{z}\,)$$

where $\Omega_j(\boldsymbol{A})$ is a matrix polynomial of degree j, usually acting as a contraction. The vector $\boldsymbol{z}$ is perpendicular to $\widehat{\boldsymbol{r}}_{n-j}$, and *is the solution of an* $(n-j)\times(n-j)$ *linear system with random matrix and random right-hand side.* This phenomenon is related to the fact that the IDR(s) procedure uses s auxiliary, randomly chosen, so-called 'shadow vectors'. These shadow vectors may be chosen real or complex. The experiments suggested that $\|\boldsymbol{z}\|$ behaves like $C\sqrt{n-j}$.

Since GMRES produces in each step the approximation $\widehat{\boldsymbol{x}}_n$ for which the residual $\widehat{\boldsymbol{r}}_n$ has minimal norm, estimates for $\|\boldsymbol{z}\|$ measure the quality of the IDR(s) algorithms.

*Delft University of Technology, Delft Institute of Applied Mathematics, Mekelweg 4, 2628 CD, The Netherlands. E-mail: p.sonneveld@tudelft.nl

2 Samples of $\|\boldsymbol{A}^{-1}\boldsymbol{b}\|$ for random $\boldsymbol{A}$ and $\boldsymbol{b}$.

We started with an experimental exploration of the distributions of $\|\boldsymbol{x}\|$, where $\boldsymbol{x}$ is the solution of an $n \times n$ system $\boldsymbol{Ax} = \boldsymbol{b}$ with Gaussian $\boldsymbol{A}$ and $\boldsymbol{b}$. The class of these stochastic vectors is denoted by $\mathcal{Q}_1^n$ if $\boldsymbol{A}$ and $\boldsymbol{b}$ are real Gaussian, and by $\mathcal{Q}_2^n$ if $\boldsymbol{A}$ and $\boldsymbol{b}$ are complex (See section 3).

We made 500 samples of this stochastic variable, for $n = 25, 50, 100, 200$, using the function 'randn' in matlab.

In the IDR(s) convergence analysis, this stochastic variable can be interpreted as a relative deviation from the GMRES convergence. Therefore we plotted the histograms of $\log_{10}(\|\boldsymbol{x}\|)$, since this quantity measures the number of digits that IDR(s) is 'behind' the GMRES procedure. The results are shown in Figure 1 and Figure 3. The histograms for different n are plotted in different colors.

A slight shift to the right can be seen at increasing n. The heuristically expected behaviour $\|\boldsymbol{x}\| \approx C\sqrt{n}$ for an $n \times n$ completely random system, would imply the shift to be $\log_{10}(n)/2$. Therefore we also plotted the histograms for $f_{\|\mathcal{Q}_\kappa^n\|}$ shifted to the left with $\log_{10}(n)/2$ in Figure 2 and Figure 4.

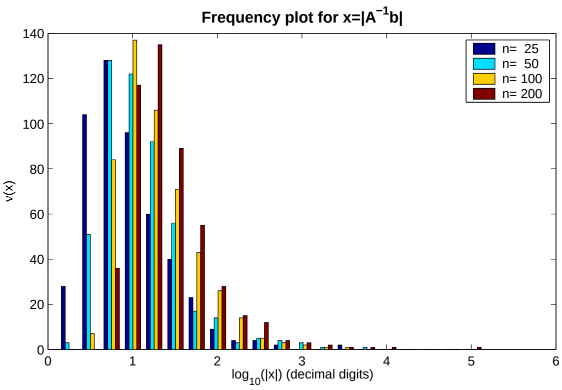

Figure 1: $\log_{10}(\|\boldsymbol{x}\|)$, $\boldsymbol{x} \in \mathcal{Q}_1^n$

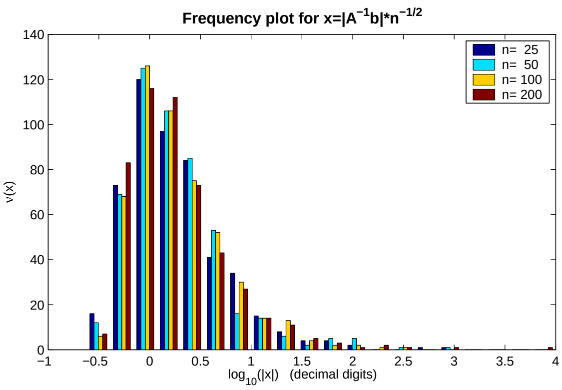

Figure 2: $\log_{10}(\frac{\|\boldsymbol{x}\|}{\sqrt{n}})$, $\boldsymbol{x} \in \mathcal{Q}_1^n$

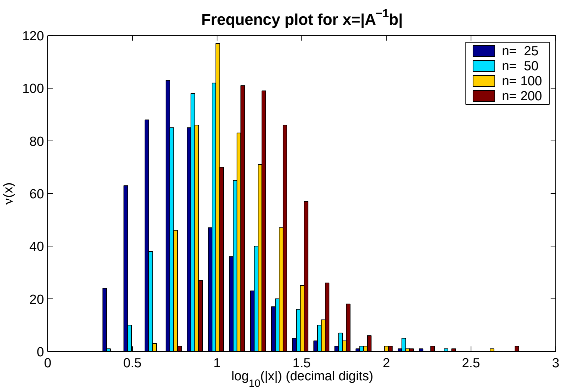

Figure 3: $\log_{10}(\|\boldsymbol{x}\|)$, $\boldsymbol{x} \in \mathcal{Q}_2^n$

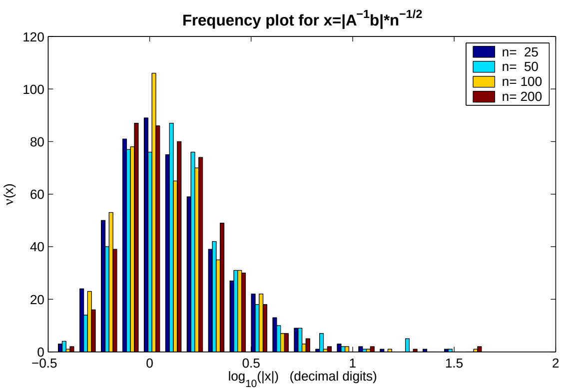

Figure 4: $\log_{10}(\frac{\|\boldsymbol{x}\|}{\sqrt{n}})$, $\boldsymbol{x} \in \mathcal{Q}_2^n$

The calculated means and variances are shown in Table 1–4.

The histograms as well as the calculated means seem to show that the variable $\log_{10}(\|\boldsymbol{x}\|/\sqrt{n})$ has a distribution function that is nearly independent of n.

n	mean	var	stdd
25	0.953	0.239	0.488
50	1.099	0.226	0.476
100	1.270	0.211	0.459
200	1.413	0.264	0.514

Table 1: $\log_{10}(\|\boldsymbol{x}\|)$, $\boldsymbol{x} \in \mathcal{Q}_1^n$

n	mean	var	stdd
25	0.254	0.239	0.488
50	0.249	0.226	0.476
100	0.270	0.211	0.459
200	0.263	0.264	0.514

Table 2: $\log_{10}(\|\boldsymbol{x}\|/\sqrt{n})$, $\boldsymbol{x} \in \mathcal{Q}_1^n$

n	mean	var	stdd
25	0.816	0.080	0.283
50	0.996	0.083	0.288
100	1.097	0.064	0.253
200	1.270	0.068	0.261

Table 3: $\log_{10}(\|\boldsymbol{x}\|)$, $\boldsymbol{x} \in \mathcal{Q}_2^n$

n	mean	var	stdd
25	0.117	0.080	0.283
50	0.146	0.083	0.288
100	0.097	0.064	0.253
200	0.120	0.068	0.261

Table 4: $\log_{10}(\|\boldsymbol{x}\|/\sqrt{n})$, $\boldsymbol{x} \in \mathcal{Q}_2^n$

We also tried to estimate $\|\boldsymbol{x}\|$ analytically, with help of available theory on random matrices, as described in [1], [2], [3] and [7]. These attempts didn't even produce a clear explanation of the $\sqrt{n}$ behaviour which is visible in the histograms. So we started searching for a better heuristic argument for this behaviour, resulting finally in a complete probability analysis for the stochastic vectors $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$, with Gaussian $\boldsymbol{A}$ and $\boldsymbol{b}$, and their norms.

3 Elementary prerequisites.

3.1 Notational conventions.

Volume element. Let $\overline{\boldsymbol{x}}$ be a stochastic variable in $\mathbb{R}^N$, with a probability density function (PDF) $f(\boldsymbol{x})$, then

$$\text{Prob}\left\{\bigcap_{k=1}^{N}\left(\ \overline{x}_k \in [x_k, x_k + dx_k)\ \right)\right\} = f(\boldsymbol{x})dx_1dx_2\cdots dx_N$$

In the case of a complex stochastic vector $\overline{\boldsymbol{z}} = \overline{\boldsymbol{x}} + i\overline{\boldsymbol{y}}$ with $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{y}}$ in $\mathbb{R}^N$, we have

$$\text{Prob}\left\{\bigcap_{k=1}^{N}\Big(\ \overline{x}_k \in [x_k, x_k + dx_k)\bigcap(\overline{y}_k \in [y_k, y_k + dy_k)\ \Big)\right\} = f(\boldsymbol{z})dx_1dy_1dx_2dy_2\cdots dx_Ndy_N$$

The difference between real and complex stochastic variables causes an essential difference in the formulae encountered. In order to improve the readability of the analysis, we'll use the following convention for volume elements like $dx_1dx_2\cdots dx_N$:

$$\boldsymbol{x} \in \mathbb{R}^N: \quad \mu(d\boldsymbol{x}) = dx_1dx_2\cdots dx_N \tag{1}$$

$$\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y} \in \mathbb{C}^N: \quad \mu(d\boldsymbol{z}) = dx_1dy_1dx_2dy_2\cdots dx_Ndy_N \tag{2}$$

Normalization constant. In working with probability densities, we often encounter formulae like

$$f(\boldsymbol{x}) = C\cdot\Phi(\boldsymbol{x}), \quad \text{with } C \text{ not depending on } \boldsymbol{x}$$

The constant C must be chosen such that

$$\iiint \cdots \int f(\boldsymbol{x})\mu(d\boldsymbol{x}) = 1$$

In many cases, the scaling constant is a beautiful, yet complicated expression, that only depends on the size of the problem.
In order to keep the formulae clear, we use the symbol $\mathfrak{C}$ in these cases. $\mathfrak{C}$ only means: *there has to be a normalization constant at this place.* So in each formula, $\mathfrak{C}$ may mean something completely different.
Only if absolutely necessary, we'll produce actual scaling constants explicitly.

3.2 Complex random matrices

As objects in probability theory, complex stochastic variables are similar to stochastic variables in $\mathbb{R}^2$. Similarly, stochastic vectors in $\mathbb{C}^n$ are similar to stochastic vectors in $\mathbb{R}^{2n}$. Every $\boldsymbol{z} \in \mathbb{C}^n$ can be represented as $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$, with $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Similarly a matrix $\boldsymbol{C} \in \mathbb{C}^{m\times n}$ can be written as $\boldsymbol{C} = \boldsymbol{A} + i\boldsymbol{B}$, with $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m\times n}$. The product $\boldsymbol{w} = \boldsymbol{C}\boldsymbol{z}$ can be calculated as

$$\boldsymbol{w} = (\boldsymbol{A} + i\boldsymbol{B})(\boldsymbol{x} + i\boldsymbol{y}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{B}\boldsymbol{y} + i(\boldsymbol{B}\boldsymbol{x} + \boldsymbol{A}\boldsymbol{y})$$

Writing $\boldsymbol{w} = \boldsymbol{u} + i\boldsymbol{v}$, with $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^m$, this product can be written as

$$\begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}$$

Every vector $\boldsymbol{z} \in \mathbb{C}^n$ is associated to a vector $\widehat{\boldsymbol{z}} \in \mathbb{R}^{2n}$, and every matrix $\boldsymbol{C} \in \mathbb{C}^{m\times n}$ is associated to a matrix $\widehat{\boldsymbol{C}} \in \mathbb{R}^{2m\times 2n}$ according to

$$\widehat{\boldsymbol{z}} = \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{y} \end{bmatrix}, \; \widehat{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{A} & -\boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{A} \end{bmatrix},$$

Then $\boldsymbol{w} = \boldsymbol{C}\boldsymbol{z} \Longleftrightarrow \widehat{\boldsymbol{w}} = \widehat{\boldsymbol{C}}\widehat{\boldsymbol{z}}$.
For the norm of a vector we have $\|\boldsymbol{z}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 = \|\widehat{\boldsymbol{z}}\|^2$. A unitary mapping $\boldsymbol{Q}$ from $\mathbb{C}^n$ to $\mathbb{C}^n$ then automatically translates into a real orthogonal mapping $\widehat{\boldsymbol{Q}}$ from $\mathbb{R}^{2n}$ to $\mathbb{R}^{2n}$.

3.3 Special mathematical formulae.

As usual, the derivation of probability density functions requires a heavy use of high-dimensional calculus, Γ-functions, and some elementary standard distributions. The relevant properties will be listed without proof; backgrounds can be found in [8].

- **Gamma function:**

  $$\Gamma(z) = \int_0^\infty \exp(-t)t^{z-1}dt, \quad \Re(z) > 0 \tag{3}$$

  The Gamma function is the only logarithmic convex solution of the functional equation

  $$\Gamma(z+1) = z\Gamma(z), \quad \Gamma(1) = 1 \tag{4}$$

Logarithm of Gamma function:

$$\Phi(z) = \log(\Gamma(z)) = -\gamma z - \log(z) - \sum_{n=1}^{\infty} \left[\log(1+\frac{z}{n}) - \frac{z}{n} \right], \quad z \neq 0, -1, -2, \ldots \quad (5)$$

where γ is Euler's constant:

$$\gamma = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k} - \log(n) \quad (6)$$

Stirlings aymptotic formula:

$$\Gamma(z) \approx z^{z-\frac{1}{2}} e^{-z} \sqrt{2\pi}(1 + O(\frac{1}{z})), \quad \text{as } z \to \infty \quad (7)$$

- **Beta function:**

$$B(a,b) = \int_0^1 u^{a-1}(1-u)^{b-1} du = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \quad (8)$$

First alternative, using $u = \sin^2(\phi)$:

$$B(a,b) = 2\int_0^{\frac{1}{2}\pi} \sin^{2a-1}(\phi) \cos^{2b-1}(\phi) d\phi \quad (9)$$

Second alternative, using $\tan(\phi) = t$:

$$B(a,b) = 2\int_0^{\infty} \frac{t^{2a-1}}{(1+t^2)^{b+a}} dt \quad (10)$$

3.4 Prerequisites from probability theory.

- **Normal distribution** $\mathcal{N}(\mu,\sigma)$, expectation $= \mu$, standard deviation $= \sigma$:

$$f_{\mathcal{N}(\mu,\sigma)}(x) = C_{\mathcal{N}(\mu,\sigma)} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \text{ with } C_{\mathcal{N}(\mu,\sigma)} = \frac{1}{\sigma\sqrt{2\pi}} \quad (11)$$

- **Standard Normal distribution** $\mathcal{N} = \mathcal{N}(0,1)$:

$$f_{\mathcal{N}}(x) = C_{\mathcal{N}} \exp\left(-\frac{x^2}{2}\right), \text{ with } C_{\mathcal{N}} = \frac{1}{\sqrt{2\pi}} \quad (12)$$

- **Real Gaussian matrices** $\mathcal{N}^{m\times n}$.

$$f_{\mathcal{N}^{m\times n}}(\boldsymbol{X}) = (C_{\mathcal{N}})^{mn} \prod_{k=1}^{m}\prod_{l=1}^{n} \exp\left(-\frac{1}{2}x_{kl}^2\right) = (C_{\mathcal{N}})^{mn} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{X}^T\boldsymbol{X})\right) \quad (13)$$

From the formula in the middle follows immediately that all entries are stochastically independent.

- χ_n **distribution**, the distribution of $\|\boldsymbol{x}\|$ when $\boldsymbol{x} \in \mathcal{N}^n$.

$$f_{\chi_n} = C_{\chi_n} \exp\left(-\frac{1}{2}x^2\right) x^{n-1}, \quad \text{with } C_{\chi_n} = \frac{1}{2^{\frac{1}{2}n-1}\Gamma(\frac{1}{2}n)} \quad (14)$$

- **Linear combinations of Gaussian variables:** Let $\boldsymbol{x} \in \mathcal{N}^n$, let $\boldsymbol{c} \in \mathbb{R}^n$, then the stochastic variable $\xi = \boldsymbol{c}^T \boldsymbol{x}$ is distributed $\mathcal{N}(0, \|\boldsymbol{c}\|)$. Therefore ξ can be written as

  $$\xi = \|\boldsymbol{c}\| z, \text{ where } z \in \mathcal{N}$$

- **Orthogonal transformation of Gaussian matrices:** Let $\boldsymbol{A} \in \mathcal{N}^{m \times n}$, let $\boldsymbol{Q}$ be a real unitary $m \times n$ matrix, and let $\boldsymbol{A}' = \boldsymbol{Q}\boldsymbol{A}$, then $\boldsymbol{A}' \in \mathcal{N}^{m \times n}$

  This follows from $(\boldsymbol{Q}\boldsymbol{A})^T(\boldsymbol{Q}\boldsymbol{A}) = \boldsymbol{A}^T(\boldsymbol{Q}^T\boldsymbol{Q})\boldsymbol{A} = \boldsymbol{A}^T\boldsymbol{A}$, and from the invariance of the volume element.

Complex Gaussian matrices. The class of complex Gaussian $m \times n$ matrices is denoted by $\widehat{\mathcal{N}}^{m \times n}$. Let $\boldsymbol{Z} \in \widehat{\mathcal{N}}^{m \times n}$, then $\boldsymbol{Z}$ has the following probability density:

$$f_{\widehat{\mathcal{N}}^{m \times n}}(\boldsymbol{Z}) = (C_{\widehat{\mathcal{N}}})^{mn} \prod_{k=1}^{m} \prod_{l=1}^{n} \exp\left(-\frac{1}{2}|z_{kl}|^2\right) = (C_{\mathcal{N}})^{2mn} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{Z}^*\boldsymbol{Z})\right) \tag{15}$$

where $\boldsymbol{Z}^* = \overline{\boldsymbol{Z}}^T$, the complex transpose. In fact this is the distribution of $2mn$ independent, normally distributed real numbers.

Generic probability densities. It is convenient to combine the analysis for real matrices and vectors. This is done by adding an extra index κ to a class identifier, where $\kappa = 1$ means real variables, and $\kappa = 2$ means complex variables. For instance

$$\mathcal{N}_\kappa^n = \begin{cases} \mathcal{N}^n & \text{if } \kappa = 1 \\ \widehat{\mathcal{N}}^n & \text{if } \kappa = 2 \end{cases}$$

Using this convention, we simply can write

$$f_{\mathcal{N}_\kappa^{m \times n}}(\boldsymbol{Z}) = (C_{\mathcal{N}})^{\kappa mn} \exp\left(-\frac{1}{2}\mathrm{tr}(\boldsymbol{Z}^*\boldsymbol{Z})\right) \tag{16}$$

Gaussian vectors are isotropic in space, i.e. the probability densities are invariant with respect to unitary transformations. In fact, the PDF only depends on the norm of the stochastic vector. This property plays an explicit role in the current analysis.
Let f be the probability density of an isotropic stochastic vector $\boldsymbol{x} \in \mathbb{R}^n$, then the marginal probability density $\widehat{f}$ with respect to $r = \|\boldsymbol{x}\|$ can be found from a 'surface integral':
Let $V_n(r)$ and $S_n(r)$ denote respectively the volume and the surface area of the sphere of radius r in $\mathbb{R}^n$, then

$$V_n(r) = \iiint \cdots \int_{\|\mathbf{x}\| < r} \mu(d\boldsymbol{x}) = r^n \iiint \cdots \int_{\|\widetilde{\mathbf{x}}\| < 1} \mu(d\widetilde{\boldsymbol{x}}) = r^n V_1(1)$$

Also we may write

$$V_n(r) = \int_0^r S_n(r)dr$$

It follows $S_n(r) = V_n'(r) = nr^{n-1}V_1(1) = r^{n-1}S_n(1)$.
The functions V_n and S_n may be calculated recursively from relations like

$$V_n(r) = 2\int_0^r V_{n-1}(\sqrt{r^2 - t^2})dtl = 2\int_0^r V_{n-1}(\widetilde{t})\frac{\widetilde{t}d\widetilde{t}}{\sqrt{r^2 - \widetilde{t}^2}}$$

but easier is using two alternative calculations for one specific integral. We use the integral

$$
\begin{aligned}
I &= \iint \cdots \int_{-\infty}^{\infty} \exp\left(-\sum_{k=1}^{n} x_k^2\right) \mu(d\boldsymbol{x}) = \pi^{\frac{1}{2}n} \\
I &= \int_0^{\infty} \exp(-r^2) S_n(r) dr = S_n(1) \int_0^{\infty} \exp(-r^2) r^{n-1} dr = \frac{1}{2} S_n(1) \Gamma(\frac{1}{2}n)
\end{aligned}
$$

from which follows

$$
S_n(1) = \frac{2\pi^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}n)} \tag{17}
$$

$S_n(1)$ is the surface area of the unit sphere in $\mathbb{R}^n$. We'll denote it simply by S_n.
Returning to isotropic stochastic variables, we can write

$$
\iiint_{\|\boldsymbol{z}\|=r} f(\|\boldsymbol{z}\|^2) \mu(dz_1, dz_2, \ldots, dz_n) = f(r^2) r^{\kappa n-1} S_{\kappa n} \cdot dr \Longrightarrow \widehat{f}(r) = f(r^2) r^{\kappa n-1} S_{\kappa n}
$$

with $\kappa = 1$ if $\boldsymbol{z} \in \mathbb{R}^n$, and $\kappa = 2$ if $\boldsymbol{z} \in \mathbb{C}^n$.

4 Completely random linear $n \times n$ systems.

4.1 Probability distributions.

We want to derive the probability distribution or -density of the (norm of the) solution of a square system with completely random matrix and right-hand side, i.e. systems $\boldsymbol{Ax} = \boldsymbol{b}$ with $\boldsymbol{A} \in \mathcal{N}_\kappa^{n\times n}$, and $\boldsymbol{b} \in \mathcal{N}_\kappa^n$. Since the set of singular random matrices is a boundary subset of $\mathcal{N}_\kappa^{n\times n}$, the probability that $\boldsymbol{A}^{-1}$ doesn't exist is zero, and we may write the solution as $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$.
We consider the solution $\boldsymbol{x}$ as an n-dimensional generalization of the quotient of two random numbers. Therefore we name the class of stochastic solution vectors $\mathcal{Q}_\kappa^n$, from 'Quotient'.

Definition 1 *A completely random linear system is a system $\boldsymbol{Ax} = \boldsymbol{b}$, with $\boldsymbol{A} \in \mathcal{N}_\kappa^{n\times n}$ and $\boldsymbol{b} \in \mathcal{N}_\kappa^n$. The class $\mathcal{Q}_\kappa^n$ is the set of solutions of completely random $n \times n$ systems:*

$$
\mathcal{Q}_\kappa^n = \left\{\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b},\ \boldsymbol{A} \in \mathcal{N}_\kappa^{n\times n},\, \boldsymbol{b} \in \mathcal{N}_\kappa^n\right\}
$$

The probability density function of a vector $\boldsymbol{x} \in \mathcal{Q}_\kappa^n$ is denoted by $f_{\mathcal{Q}_\kappa^n}(\boldsymbol{x})$. The PDF of the norm $\|\boldsymbol{x}\|$ for these vectors is denoted by $f_{\|\mathcal{Q}_\kappa^n\|}(x)$, where $x = \|\boldsymbol{x}\|$.

In the derivation of the PDF 's we need the following lemma

Lemma 1 *Let $\alpha \sim \chi_n$, for $n \geq 1$ and $\boldsymbol{b} \sim \mathcal{N}^k$, $k \geq 1$, be stochastically independent. Let $\boldsymbol{x} = \dfrac{\boldsymbol{b}}{\alpha}$, then this stochastic vector has the PDF*

$$
f(\boldsymbol{x}) = \frac{C_k^n}{(1 + \|\boldsymbol{x}\|^2)^{\frac{1}{2}(n+k)}} \tag{18}
$$

with

$$
C_k^n = \frac{\Gamma(\frac{1}{2}(n+k))}{\pi^{\frac{k}{2}}\Gamma(\frac{n}{2})} \tag{19}
$$

Proof: Let $g(\alpha, \boldsymbol{b})$ be the joint distribution of α and $\boldsymbol{b}$, then since α and $\boldsymbol{b}$ are stochastically independent, g is the product of the PDF 's for α and $\boldsymbol{b}$:

$$g(\alpha, \boldsymbol{b}) = \mathfrak{C} \exp(-\frac{1}{2}(\alpha^2 + \|\boldsymbol{b}\|^2))\alpha^{n-1}$$

Define the new stochastic variables s and $x_1, x_2, \ldots, x_k$:

$$\alpha = s, \ b_j = sx_j, \ j = 1, 2, \ldots, k$$

Then the Jacobian of this transformation in block form reads

$$\boldsymbol{J} = \frac{\partial(\alpha, b_1, b_2, \ldots, b_k)}{\partial(s, x_1, x_2, \ldots, x_k)} = \begin{bmatrix} 1 & \boldsymbol{0}^T \\ \boldsymbol{x} & s\boldsymbol{I} \end{bmatrix} \Longrightarrow \det(\boldsymbol{J}) = s^k$$

and we have

$$\mu(d\alpha d\boldsymbol{b}) = |\det(\boldsymbol{J})| \, \mu(ds d\boldsymbol{x}) = s^k ds \mu(d\boldsymbol{x})$$

Furthermore:

$$\alpha^2 + \|\boldsymbol{b}\|^2 = s^2(1 + \|\boldsymbol{x}\|^2)$$

Denote the mutual density for $s, \boldsymbol{x}$ by $\widehat{g}(s, \boldsymbol{x})$, then $g(\alpha, \boldsymbol{b})\mu(d\alpha d\boldsymbol{b}) = \widehat{g}(s, \boldsymbol{x})\mu(ds d\boldsymbol{x})$, and therefore

$$\widehat{g}(s, \boldsymbol{x}) = g(\alpha, \boldsymbol{b}) \cdot |\det(\boldsymbol{J})| = \mathfrak{C} \exp\left(-\frac{1}{2}s^2(1 + \|\boldsymbol{x}\|^2)\right) s^{n+k-1}$$

The PDF $f(\boldsymbol{x})$ is the marginal distribution of $\boldsymbol{x}$, i.e. the integral

$$\begin{aligned} f(\boldsymbol{x}) &= \int_0^\infty \widehat{g}(s, \boldsymbol{x}) ds = \mathfrak{C} \int_0^\infty \exp(-\frac{1}{2}s^2(1 + \|\boldsymbol{x}\|^2))s^{n+k-1} \\ &= \mathfrak{C} \int_0^\infty \frac{\exp(-\frac{1}{2}t^2)t^{n+k-1}}{(1 + \|\boldsymbol{x}\|^2)^{\frac{1}{2}(n+k)}} dt \\ &= \frac{\mathfrak{C}'}{(1 + \|\boldsymbol{x}\|^2)^{\frac{1}{2}(n+k)}} \end{aligned}$$

The actual value C_k^n for $\mathfrak{C}'$ must satisfy

$$[C_k^n]^{-1} = \iint \cdots \int \frac{\mu(d\boldsymbol{x})}{(1 + \|\boldsymbol{x}\|^2)^{\frac{1}{2}(n+k)}} = S_k \int_0^\infty \frac{r^{k-1} dr}{(1 + r^2)^{\frac{1}{2}(n+k)}} = \frac{2\pi^{\frac{k}{2}}}{\Gamma(\frac{k}{2})} \cdot \frac{1}{2} B(a, b)$$

with $2a = k$, and $a + b = \frac{1}{2}(n + k)$, according to (10). Working this out, we get

$$C_k^n = \frac{\Gamma(\frac{k}{2})}{2\pi^{\frac{k}{2}}} \cdot \frac{2\Gamma(\frac{n+k}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{k}{2})} = \frac{\Gamma(\frac{n+k}{2})}{\pi^{\frac{k}{2}}\Gamma(\frac{n}{2})}$$

which proves the lemma. $\triangle$

Theorem 1 *The stochastic vectors in $\mathcal{Q}_\kappa^n$ have the following probability density function:*

$$f_{\mathcal{Q}_\kappa^n}(\boldsymbol{x}) = \frac{C_{\mathcal{Q}_\kappa^n}}{(1 + \|\boldsymbol{x}\|^2)^{\frac{\kappa}{2}(n+1)}}, \tag{20}$$

with

$$C_{\mathcal{Q}_\kappa^n} = \frac{\Gamma(\frac{\kappa}{2}(n + 1))}{\Gamma(\frac{\kappa}{2}) \cdot \pi^{\frac{\kappa}{2}n}} \tag{21}$$

Proof: Denote the columns of $\boldsymbol{A}$ by $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots \boldsymbol{a}_n$. Let $\boldsymbol{Q}$ be unitary, such that $\boldsymbol{QA}$ has $\|\boldsymbol{a}_n\|\boldsymbol{e}_n$ as last column. Such a $\boldsymbol{Q}$ can be constructed with for instance a Householder reflection matrix $\boldsymbol{H} = \boldsymbol{I} - 2\frac{\boldsymbol{u}\boldsymbol{u}^*}{\|\boldsymbol{u}\|^2}$, with $\boldsymbol{u} = \boldsymbol{a}_n - \|\boldsymbol{a}_n\|\boldsymbol{e}_n$. With this choice, $\boldsymbol{H}$ is a stochastic matrix *that depends only on* $\boldsymbol{a}_n$. Therefore $\boldsymbol{H}$ is stochastically independent from $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_{n-1}$ and $\boldsymbol{b}$.
Let $\boldsymbol{HA} = \boldsymbol{A}'$, $\boldsymbol{Hb} = \boldsymbol{b}'$, then the system $\boldsymbol{Ax} = \boldsymbol{b}$ is equivalent to the system $\boldsymbol{A}'\boldsymbol{x} = \boldsymbol{b}'$. Moreover, all entries of $\boldsymbol{A}'$ except those in the last column, and all entries of $\boldsymbol{b}'$ are stochastically independent, and standard normally distributed. (prerequisite 3.4).
We can write the matrix and the relevant vectors in blocks:

$$\boldsymbol{x} = \begin{bmatrix} \widetilde{\boldsymbol{x}} \\ x_n \end{bmatrix}, \ \boldsymbol{b}' = \begin{bmatrix} \widetilde{\boldsymbol{b}} \\ b'_n \end{bmatrix}, \ \boldsymbol{A}' = \begin{bmatrix} \widetilde{\boldsymbol{A}} & \boldsymbol{0} \\ \widetilde{\boldsymbol{a}}^* & \alpha \end{bmatrix}$$

with $\alpha = a'_{nn} = \|\boldsymbol{a}_n\|$. The system can now be written as

$$\alpha x_n = b'_n - \widetilde{\boldsymbol{a}}^*\widetilde{\boldsymbol{x}}, \ \text{ with } \widetilde{\boldsymbol{A}}\widetilde{\boldsymbol{x}} = \widetilde{\boldsymbol{b}}$$

Here $\widetilde{\boldsymbol{x}}$ is completely determined by $\widetilde{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{b}}$. Now $\widetilde{\boldsymbol{A}} \in \mathcal{N}_\kappa^{(n-1)\times(n-1)}$, $\widetilde{\boldsymbol{b}} \in \mathcal{N}_\kappa^{n-1}$, and hence $\widetilde{\boldsymbol{x}} \in \mathcal{Q}_\kappa^{n-1}$. Furthermore $\widetilde{\boldsymbol{a}} \in \mathcal{N}_\kappa^{n-1}$, $b'_n \in \mathcal{N}_\kappa$, and $\alpha \in \chi_{\kappa n}$. Finally $\widetilde{\boldsymbol{a}}, \alpha$ and b'_n are stochastically independent from $\widetilde{\boldsymbol{x}}$, since these variables are stochastically independent from $\widetilde{\boldsymbol{A}}$ and $\widetilde{\boldsymbol{b}}$.

We are after the PDF of $\boldsymbol{x}$, that is the joint distribution of $\widetilde{\boldsymbol{x}}$ and x_n. This is the product of $f_{\mathcal{Q}_\kappa^{n-1}}(\widetilde{\boldsymbol{x}})$ — the PDF of $\widetilde{\boldsymbol{x}}$— and the distribution of x_n *conditional with respect to* $\widetilde{\boldsymbol{x}}$:

$$f_{\mathcal{Q}_\kappa^n}(\boldsymbol{x}) = f_{\mathcal{Q}_\kappa^n}(\widetilde{\boldsymbol{x}}, x_n) = f_{\mathcal{Q}_\kappa^n}(\widetilde{\boldsymbol{x}}, x_n \,|\, \widetilde{\boldsymbol{x}}) \cdot f_{\mathcal{Q}_\kappa^{n-1}}(\widetilde{\boldsymbol{x}})$$

We'll denote the conditional density by $g(x_n)$:

$$g(x_n) = f_{\mathcal{Q}_\kappa^n}(\widetilde{\boldsymbol{x}}, x_n \,|\, \widetilde{\boldsymbol{x}}) = \frac{f_{\mathcal{Q}_\kappa^n}(\widetilde{\boldsymbol{x}}, x_n)}{f_{\mathcal{Q}_\kappa^{n-1}}(\widetilde{\boldsymbol{x}})} = \frac{f_{\mathcal{Q}_\kappa^n}(\boldsymbol{x})}{f_{\mathcal{Q}_\kappa^{n-1}}(\widetilde{\boldsymbol{x}})} \tag{22}$$

g is the PDF of the stochastic quotient

$$x_n = \frac{b'_n - \widetilde{\boldsymbol{a}}^*\widetilde{\boldsymbol{x}}}{\alpha} = \frac{\beta}{\alpha}$$

in which $\widetilde{\boldsymbol{x}}$ is considered a known vector. The numerator β of this expression is a linear combination of n stochastically independent standard normally distributed variables:

$$\beta = b'_n - \widetilde{\boldsymbol{a}}^*\widetilde{\boldsymbol{x}} = [\widetilde{\boldsymbol{x}}^* \ 1] \begin{bmatrix} -\widetilde{\boldsymbol{a}} \\ b'_n \end{bmatrix}$$

and according to prerequisite 3.4, we have $\beta = \sqrt{1 + \|\widetilde{\boldsymbol{x}}\|^2} \cdot w$, with $w \in \mathcal{N}_\kappa$. Now let $x_n = \sqrt{1 + \|\widetilde{\boldsymbol{x}}\|^2}\xi$, then

$$\xi = \frac{x_n}{\sqrt{1 + \|\widetilde{\boldsymbol{x}}\|^2}} = \frac{w}{\alpha} \tag{23}$$

with $w \in \mathcal{N}_\kappa$, and $\alpha \in \chi_{\kappa n}$. Bearing in mind the equivalence of $\mathcal{N}_\kappa$ and $\mathcal{N}^\kappa$, we may apply lemma 1 with $k = \kappa$.

$$f(\xi) = \frac{C_\kappa^{\kappa n}}{(1 + \xi^2)^{\frac{\kappa}{2}(n+1)}} \tag{24}$$

$f(\xi)$ and $g(x_n)$ are related by $g(x_n)\mu(dx_n) = f(\xi)\mu(d\xi)$. Using $\mu(dx_n) = \left(\sqrt{1+\|\widetilde{\boldsymbol{x}}\|^2}\right)^\kappa \mu(d\xi)$, we can write

$$g(x_n)\mu(dx_n) = f(\xi)\mu(d\xi) \Longrightarrow g(x_n) = f(\xi)\frac{\mu(d\xi)}{\mu(dx_n)} = \frac{f(\xi)}{(1+\|\widetilde{\boldsymbol{x}}\|^2)^{\frac{\kappa}{2}}} \tag{25}$$

Now since

$$\frac{1}{1+|\xi|^2} = \left(1+\frac{|x_n|^2}{1+\|\widetilde{\boldsymbol{x}}\|^2}\right)^{-1} = \frac{1+\|\widetilde{\boldsymbol{x}}\|^2}{1+|x_n|^2+\|\widetilde{\boldsymbol{x}}\|^2} = \frac{1+\|\widetilde{\boldsymbol{x}}\|^2}{1+\|\boldsymbol{x}\|^2}$$

(24) can be written as

$$f(\xi) = C_\kappa^{\kappa n}\left(\frac{1+\|\widetilde{\boldsymbol{x}}\|^2}{1+\|\boldsymbol{x}\|^2}\right)^{\frac{\kappa}{2}(n+1)}$$

Substitution this in (25) then produces the PDF $g(x_n)$ explicitly:

$$g(x_n) = \frac{f(\xi)}{(1+\|\widetilde{\boldsymbol{x}}\|^2)^{\frac{\kappa}{2}}} = C_\kappa^{\kappa n}\frac{(1+\|\widetilde{\boldsymbol{x}}\|^2)^{\frac{\kappa}{2}n}}{(1+|x_n|^2+\|\widetilde{\boldsymbol{x}}\|^2)^{\frac{\kappa}{2}(n+1)}} \tag{26}$$

We now prove the theorem by induction. The induction hypothesis is (20).
If $n=1$, the conditional density is the absolute density, since there is no $\widetilde{\boldsymbol{x}}$. So $\xi = x_n = x$, the only entry of the vector $\boldsymbol{x}$. Therefore we have

$$f_{\mathcal{Q}_\kappa^1}(x) = \frac{C_{\mathcal{Q}_\kappa^1}}{(1+|x|^2)^\kappa} = \frac{C_\kappa^\kappa}{(1+|x|^2)^\kappa} = \frac{1}{\pi}\cdot\frac{1}{(1+|x|^2)^\kappa} \tag{27}$$

Here the value π^{-1} for C_1^1 as well for C_2^2 follows directly from (19). By comparing (21) with $n=1$ and (19) with $k=n=\kappa$, we find $C_{\mathcal{Q}_\kappa^1} = C_\kappa^\kappa = \pi^{-1}$ for $\kappa = 1, 2$. Finally, it is easily verified that

$$\int_{\mathbb{R}^\kappa}\frac{1}{\pi}\cdot\frac{1}{(1+|x|^2)^\kappa}\mu(dx) = 1$$

According to (22), and the induction hypothesis for $n-1$, we must have

$$\begin{aligned} f_{\mathcal{Q}_\kappa^n}(\boldsymbol{x}) &= g(x_n)f_{\mathcal{Q}_\kappa^{n-1}}(\widetilde{\boldsymbol{x}}) = C_\kappa^{\kappa n}\frac{(1+\|\widetilde{\boldsymbol{x}}\|^2)^{\frac{\kappa}{2}n}}{(1+|x_n|^2+\|\widetilde{\boldsymbol{x}}\|^2)^{\frac{\kappa}{2}(n+1)}}\cdot\frac{C_{\mathcal{Q}_\kappa^{n-1}}}{(1+\|\widetilde{\boldsymbol{x}}\|^2)^{\frac{\kappa}{2}n}} \\ &= \frac{C_\kappa^{\kappa n}C_{\mathcal{Q}_\kappa^{n-1}}}{(1+\|\boldsymbol{x}\|^2)^{\frac{\kappa}{2}(n+1)}} \end{aligned}$$

This is (20), provided the constants satify $C_{\mathcal{Q}_\kappa^n} = C_\kappa^{\kappa n}C_{\mathcal{Q}_\kappa^{n-1}}$. This is easily verified:

$$C_\kappa^{\kappa n}C_{\mathcal{Q}_\kappa^{n-1}} = \frac{\Gamma(\frac{\kappa}{2}(n+1))}{\pi^{\frac{\kappa}{2}}\Gamma(\frac{\kappa}{2}n)}\cdot\frac{\Gamma(\frac{\kappa}{2}n)}{\pi^{\frac{\kappa(n-1)}{2}}\Gamma(\frac{\kappa}{2})} = \frac{\Gamma(\frac{\kappa}{2}(n+1))}{\pi^{\frac{\kappa n}{2}}\Gamma(\frac{\kappa}{2})}$$

which is $C_{\mathcal{Q}_\kappa^n}$ according to (21). Furthermore, the constant $C_{\mathcal{Q}_\kappa^n}$ must satisfy

$$\left(C_{\mathcal{Q}_\kappa^n}\right)^{-1} = \int\frac{\mu(d\boldsymbol{x})}{(1+\|\boldsymbol{x}\|^2)^{\frac{\kappa}{2}(n+1)}}$$

Using $\int_{\|\mathbf{x}\|=1} \mu(d\boldsymbol{x}) = S_{\kappa n} r^{\kappa n-1} dr$ and (10), this results in

$$\left(C_{\mathcal{Q}_\kappa^n}\right)^{-1} = S_{\kappa n} \cdot \int_0^\infty \frac{r^{\kappa n-1} dr}{(1+r^2)^{\frac{\kappa}{2}(n+1)}} = \frac{2\pi^{\frac{\kappa n}{2}}}{\Gamma(\frac{\kappa n}{2})} \cdot \frac{1}{2} B(a,b)$$

with $2a-1 = \kappa n - 1$, $a+b = \frac{\kappa}{2}(n+1)$. It follows

$$C_{\mathcal{Q}_\kappa^n} = \frac{\Gamma(\frac{\kappa n}{2})}{\pi^{\frac{\kappa n}{2}}} \cdot \frac{\Gamma(\frac{\kappa}{2}(n+1))}{\Gamma(\frac{\kappa}{2}n)\Gamma(\frac{\kappa}{2})} = \frac{\Gamma(\frac{\kappa}{2}(n+1))}{\pi^{\frac{\kappa n}{2}}\Gamma(\frac{\kappa}{2})}$$

which is (21). This proves the theorem. △

The PDF for vectors $\boldsymbol{x}$ in $\mathcal{Q}_\kappa^n$ is symmetric in the components of $\boldsymbol{x}$, which is natural because permutations of columns of $\boldsymbol{A}$ do not change the stochastic distribution of $\boldsymbol{A}$. More interesting is the fact that the PDF only depends on $\|\boldsymbol{x}\|$. This makes is quite easy to derive the PDF for the norms of vectors in $\mathcal{Q}_\kappa^n$, which was required in the IDR(s) convergence analysis.

Theorem 2 *Let $\boldsymbol{x} \in \mathcal{Q}_\kappa^n$, let $x = \|\boldsymbol{x}\|$, then x is a stochastic variable with the following PDF :*

$$f_{\|\mathcal{Q}_\kappa^n\|}(x) = C_{\|\mathcal{Q}_\kappa^n\|} \cdot \frac{x^{\kappa n-1}}{(1+x^2)^{\frac{\kappa}{2}(n+1)}} \tag{28}$$

with

$$C_{\|\mathcal{Q}_\kappa^n\|} = \frac{2\Gamma(\frac{\kappa}{2}(n+1)}{\Gamma(\frac{\kappa}{2}n)\Gamma(\frac{\kappa}{2})} \tag{29}$$

Proof: The required PDF can be defined as

$$\begin{aligned} f_{\|\mathcal{Q}_\kappa^n\|}(r) &= \iiint_{\|\mathbf{x}\|=r} f_{\mathcal{Q}_\kappa^n}(\boldsymbol{x}) \mu(dx_1 dx_2 \cdots dx_n) \\ &= C_{\mathcal{Q}_\kappa^n} S_{\kappa n} \frac{r^{\kappa n-1}}{(1+r^2)^{\frac{\kappa}{2}(n+1)}} = \mathfrak{C} \frac{r^{\kappa n-1}}{(1+r^2)^{\frac{\kappa}{2}(n+1)}} \end{aligned}$$

The constant $\mathfrak{C}$ must satisfy

$$\mathfrak{C}^{-1} = \int_0^\infty \frac{r^{\frac{\kappa}{2}(n-1)}}{(1+r^2)^{\frac{\kappa}{2}(n+1)}} = \frac{1}{2} B(a,b)$$

with $a = \frac{\kappa}{2}n$ and $b = \frac{\kappa}{2}$, according to (10). Hence

$$C_{\|\mathcal{Q}_\kappa^n\|} = \frac{2}{B(\frac{\kappa}{2}n, \frac{\kappa}{2})} = \frac{2\Gamma\frac{\kappa}{2}(n+1))}{\Gamma(\frac{\kappa}{2}n)\Gamma(\frac{\kappa}{2})} = C_{\mathcal{Q}_\kappa^n} S_{\kappa n}$$

This proves the theorem. △

4.2 Percentage points for $f_{\|\mathcal{Q}_\kappa^n\|}$.

For practical use, it is interesting which values of $\|\boldsymbol{x}\|$ will occur with probability less than a particular value, say 10^{-j}. It is easy to give estimates that are reasonably sharp for $j \geq 2$.
For $\boldsymbol{x} \in \mathcal{Q}_\kappa^n$, we have

$$\begin{aligned}\text{Prob}(\|\boldsymbol{x}\| > \lambda) &= C_{\|\mathcal{Q}_\kappa^n\|} \int_\lambda^\infty \frac{x^{\kappa n-1}dx}{(1+x^2)^{\frac{\kappa}{2}(n+1)}} < C_{\|\mathcal{Q}_\kappa^n\|} \int_\lambda^\infty \frac{x^{\kappa n-1}dx}{x^{\kappa(n+1)}} \\ &= C_{\|\mathcal{Q}_\kappa^n\|} \int_\lambda^\infty x^{-1-\kappa}dx = C_{\|\mathcal{Q}_\kappa^n\|}\frac{\lambda^{-\kappa}}{\kappa} \end{aligned} \tag{30}$$

In order to find the asymptotic behaviour of $C_{\|\mathcal{Q}_\kappa^n\|}$ for large n, we need the asymptotic behaviour for $\Gamma(z+\alpha)/\Gamma(z)$ for large z. With Stirlings formula (7) we may write

$$\frac{\Gamma(z+\alpha)}{\Gamma(z)} \approx \frac{(z+\alpha)^{z+\alpha-\frac{1}{2}}e^{-z-\alpha}}{z^{z-\frac{1}{2}}e^{-z}} = (z+\alpha)^\alpha e^{-\alpha}\left(1+\frac{\alpha}{z}\right)^{z-\frac{1}{2}} \approx z^\alpha(1+O(\frac{1}{z}))$$

Now, with $\alpha = \frac{\kappa}{2}$ and $\Gamma(\frac{\kappa}{2}) = \pi^{1-\frac{\kappa}{2}}$ for $\kappa = 1, 2$, we get

$$C_{\|\mathcal{Q}_\kappa^n\|} = 2\frac{\Gamma\frac{\kappa}{2}(n+1))}{\Gamma(\frac{\kappa}{2}n)\Gamma(\frac{\kappa}{2})} \approx 2\frac{\left(\frac{\kappa}{2}n\right)^{\frac{\kappa}{2}}}{\Gamma(\frac{\kappa}{2})} = \frac{2}{\pi}(\frac{\kappa}{2}\pi n)^{\frac{\kappa}{2}} \tag{31}$$

Hence, with (30), we arrive at

$$\text{Prob}(\|\boldsymbol{x}\| > \lambda) < C_{\|\mathcal{Q}_\kappa^n\|}\frac{\lambda^{-\kappa}}{\kappa} \approx \frac{2}{\pi}\left(\frac{\kappa}{2}\pi n\right)^{\frac{\kappa}{2}}\frac{\lambda^{-\kappa}}{\kappa} = \left(\frac{2}{\pi\kappa}\right)^{1-\frac{\kappa}{2}}\left(\frac{\sqrt{n}}{\lambda}\right)^\kappa$$

We want this probability to be smaller than 10^{-j}, and this is acomplished by choosing

$$\left(\frac{2}{\pi\kappa}\right)^{1-\frac{\kappa}{2}}\left(\frac{\sqrt{n}}{\lambda}\right)^\kappa = 10^{-j} \Longrightarrow \lambda = \sqrt{\frac{n\pi\kappa}{2}}\left(\frac{2}{\pi\kappa}10^j\right)^{\frac{1}{\kappa}}$$

For $j = 2$ we get the 99%-point for this distribution:

$$\begin{aligned}\kappa = 1 \quad &: \quad \lambda = \sqrt{\frac{2n}{\pi}}10^2 \approx 80\sqrt{n} \\ \kappa = 2 \quad &: \quad \lambda = 10\sqrt{n}\end{aligned}$$

4.3 Expectation and variance.

The distribution $f_{\|\mathcal{Q}_1^n\|}$ of $\|\boldsymbol{x}\|$ with $\boldsymbol{x} \in \mathcal{Q}_1^n$ has no finite moments, and the distribution $f_{\|\mathcal{Q}_2^n\|}$ only has a finite first moment. In the IDR(s) application, we are interested in the behaviour of $\log_{10}(\|\boldsymbol{x}\|)$ rather than $\|\boldsymbol{x}\|$ itself, since the 10-logarithm represents the number of digits that the IDR(s) process is 'behind' the GMRES procedure. Therefore we study the *logarithmic moments*.

Let $\mathfrak{C}f(x)$ be the PDF for some nonnegative stochastic variable, then the logarithmic moments are defined as

$$\mu_k = \mathfrak{C}\int_0^\infty f(x)\log^k(x)dx$$

These moments can be found by expanding the following generating function $F(t)$ in powers of t:

$$F(t) = \int_0^\infty f(x)x^t dx = \mathfrak{C}^{-1}\sum_{k=0}^\infty \mu_k \frac{t^k}{k!}$$

using $x^t = \exp(t\log(x))$. Then we have, since $\mu_0 = 1$

$$F^{(k)}(0) = \mathfrak{C}^{-1}\mu_k = F(0)\mu_k,\ \ k = 0, 1, 2, \ldots$$

It follows

$$\mu_1 = \frac{F'(0)}{F(0)}, \quad \mu_2 = \frac{F''(0)}{F(0)}, \quad \sigma^2 = \mu_2 - \mu_1^2 = \frac{F''(0)}{F(0)} - \left[\frac{F'(0)}{F(0)}\right]^2$$

For μ_1 and σ^2 this can be written elegantly as

$$\mu_1 = \frac{d}{dt}\log(F(t))|_{t=0}\,, \quad \sigma^2 = \frac{d^2}{dt^2}\log(F(t))|_{t=0} \tag{32}$$

The PDF 's related to completely random systems read

$$f(x) = \mathfrak{C}\frac{x^p}{(1+x^2)^q}$$

with $p = \kappa n - 1,\ q = \frac{\kappa}{2}(n+1)$, where $\kappa = 1$ for the real distributions, and $\kappa = 2$ for the complex distributions. Calculating the integral of $f(x)x^t$ is a mere replacement of p by $p + t$. Using (10) we get

$$F(t) = \int_0^\infty \frac{x^{p+t}}{(1+x^2)^q}dx = \frac{1}{2}B(a,b)$$

with $2a - 1 = p + t$, $a + b = q$. It follows

$$a = \frac{1}{2}(p+t+1) = \frac{1}{2}(\kappa n + t), \quad b = q - a = \frac{1}{2}(\kappa - t)$$

Hence we have

$$F(t) = \frac{1}{2}B(\frac{\kappa n + t}{2}, \frac{\kappa - t}{2}) = \frac{1}{2}\frac{\Gamma(\frac{1}{2}(\kappa n + t))\Gamma(\frac{1}{2}(\kappa - t))}{\Gamma(\frac{\kappa}{2}(n+1))}$$

Let $\log(F(t)) = \Psi(t)$, then using (5) we get

$$\Psi(t) = \Phi(\frac{\kappa n + t}{2}) + \Phi(\frac{\kappa - t}{2}) - \Phi(\frac{\kappa(n+1)}{2}) - \log(2)$$

According to (32), we now may write

$$\mu_1 = \Psi'(0) = \frac{1}{2}\left(\Phi'(\frac{\kappa n}{2}) - \Phi'(\frac{\kappa}{2})\right) \tag{33}$$

$$\sigma^2 = \Psi''(0) = \frac{1}{4}\left(\Phi''(\frac{\kappa n}{2}) + \Phi''(\frac{\kappa}{2})\right) \tag{34}$$

Differentiating (5), we find series expansions for the derivatives of the $\Phi(z)$:

$$\begin{aligned}\Phi'(z) &= -\gamma - \frac{1}{z} + \sum_{k=1}^\infty \left[\frac{1}{k} - \frac{1}{k+z}\right] \\ \Phi''(z) &= \frac{1}{z^2} + \sum_{k=1}^\infty \frac{1}{(k+z)^2}\end{aligned}$$

We need to evaluate these expressions for $z = \frac{\kappa}{2}n$, for integer values of n and for $\kappa = 1, 2$. Fortunately, $\Phi'(z)$ and $\Phi''(z)$ can be obtained in closed form for this family of z-values.

Define the (hyper-) harmonic sums by

$$S_n^{(r)} = \sum_{k=1}^{n} \frac{1}{k^r}$$

Then for integer values of l, we find after some calculations

$$\Phi'(l) = \Phi'(1) + S_{l-1}^{(1)} \quad , \quad \Phi''(l) = \frac{\pi^2}{6} - S_{l-1}^{(2)}$$

$$\Phi'(l + \frac{1}{2}) = \frac{\pi^2}{2} + 2S_{2l}^{1} - S_l^{(1)} \quad , \quad \Phi''(l + \frac{1}{2}) = \Phi''(\frac{1}{2}) - 4S_{2l}^{(2)} + S_l^{(2)}$$

Now we evaluate the expressions (33) and (34)

$$n = 2l,\ \kappa = 1: \qquad \mu_1 = \frac{1}{2}S_{l-1}^{(1)} + \log(2), \quad \sigma^2 = \frac{1}{6}\pi^2 - \frac{1}{4}S_{l-1}^{(2)}$$

$$n = 2l+1,\ \kappa = 1: \qquad \mu_1 = S_{2l}^{(1)} - \frac{1}{2}S_l^{(1)}, \quad \sigma^2 = \frac{1}{4}(\pi^2 + S_l^{(2)}) - S_{2l}^{(2)}$$

$$\kappa = 2: \qquad \mu_1 = \frac{1}{2}S_{n-1}, \quad \sigma^2 = \frac{\pi^2}{12} - \frac{1}{4}S_{n-1}^{(2)}$$

For large n we get approximately

$$\kappa = 1 \quad : \quad \mu(n) \approx \frac{1}{2}(\log(2n) + \gamma) + O(\frac{1}{n}), \quad \sigma^2(n) \approx \frac{1}{8}\pi^2 + O(\frac{1}{n}) \tag{35}$$

$$\kappa = 2 \quad : \quad \mu(n) \approx \frac{1}{2}(\log(n) \ + \gamma) + O(\frac{1}{n}), \quad \sigma^2(n) \approx \frac{1}{24}\pi^2 + O(\frac{1}{n}) \tag{36}$$

4.4 Asymptotic behaviour.

We try to find out wether this distribution, when x is scaled by $\sqrt{n}$, indeed depends only weakly on n.

$$f_{\|\mathcal{Q}_\kappa^n\|}(x) = C_{\|\mathcal{Q}_\kappa^n\|} \frac{x^{\kappa n - 1}}{(1 + x^2)^{\frac{\kappa}{2}(n+1)}}$$

For $\alpha \in (-1, 1)$, and $m > 0$, the following inequality holds ([8]):

$$(1 - m\alpha^2)e^{m\alpha} < (1 + \alpha)^m < e^{m\alpha} \Longrightarrow (1 + \alpha)^m = (1 - \theta m\alpha^2)e^{m\alpha}$$

where $\theta \in (0, 1)$. With $\alpha = x^{-2}$, and $m = \frac{\kappa}{2}(n + 1)$, we can apply this to $f_{\|\mathcal{Q}_\kappa^n\|}(x)$:

$$\frac{x^{\kappa n - 1}}{(1 + x^2)^{\frac{\kappa}{2}(n+1)}} = x^{-\kappa - 1} \exp(-\frac{\kappa(n+1)}{2x^2}) \left(1 - \theta\frac{\kappa(n+1)}{2x^4}\right)^{-1} \tag{37}$$

Now let y be defined by

$$x^2 = \frac{\kappa}{2}(n + 1)y^2$$

and let $g_n(y)$ denote the PDF for y, then $f_{\|\mathcal{Q}_\kappa^n\|}(x)dx = g_n(y)dy$. Using (31) and (37), we can write

$$g_n(y)dy = f_{\|\mathcal{Q}_\kappa^n\|}(x)dx = C_{\|\mathcal{Q}_\kappa^n\|} x^{-\kappa-1} \exp(-\frac{\kappa(n+1)}{2x^2}) \left(1 - \theta\frac{\kappa(n+1)}{2x^4}\right)^{-1} dx$$

$$\approx \pi^{\frac{\kappa}{2} - 1} \cdot y^{-\kappa-1} \exp(-\frac{1}{y^2})dy \tag{38}$$

For $y \to 0$ and fixed n, $g_n(y)$ tends to zero like $Cy^{\kappa n-1}$ for some C, whereas the asymptotic approximation is tending to zero faster than any power of y. Despite a completely different behaviour, the difference for small y is hardly visible in practice.
For large values of y, the aymptotic approximation is uniform $O\left(\frac{1}{n}\right)$. Therefore the asymptotic formula may replace the original distribution perfectly well if only n is not too small, say $n > 20$.
In Figure 6 and Figure 8 of the next section, the densities $g_n(y)$ are plotted against $\log_{10}(y)$ for $n = 25, 50, 100, 200$. Also the means and variances are displayed. The result is in very good agreement with (38) as well as with the experimental histograms in Figure 2 and Figure 4.

4.5 Numerical verification.

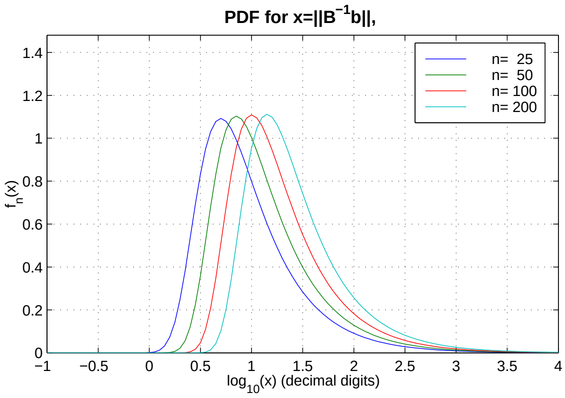

Figure 5: $f_{\|\mathcal{Q}_1^n\|}(x)$

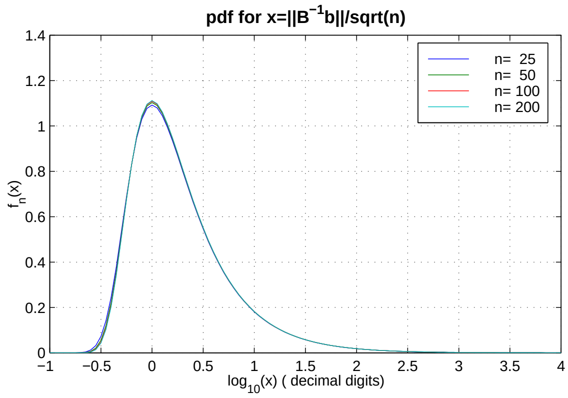

Figure 6: $f_{\|\mathcal{Q}_1^n\|}(x)$, shifted

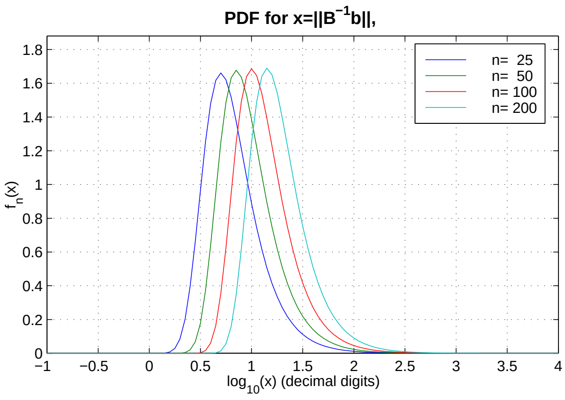

Figure 7: $f_{\|\mathcal{Q}_2^n\|}(x)$

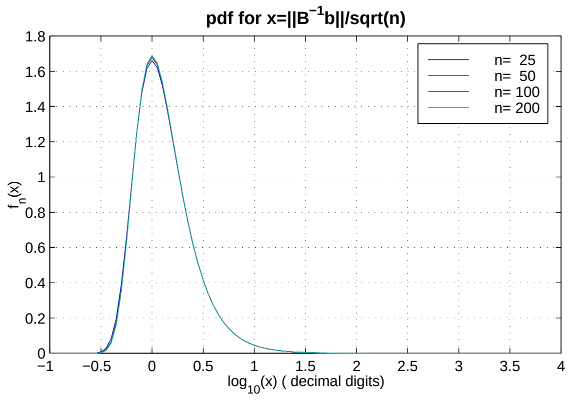

Figure 8: $f_{\|\mathcal{Q}_2^n\|}(x)$, shifted

We now verify the theory with the sampling experiments shown in section 2. We plot the analytic densities for the dimensions $n = 25, 50, 100, 200$.

In Figure 5 and Figure 7 the densities $f_{\|\mathcal{Q}_1^n\|}$ and $f_{\|\mathcal{Q}_2^n\|}$ are plotted against $\log_{10}(x)$ for $n = 25, 50, 100, 200$. The shifted variants are present in Figure 6 and Figure 8. We also calculate the mean and variance for $\log_{10}(x)$, see Table 5–8

The analytic results are in good agrement with the sampling experiments.

n	mean	var	stdd
25	0.9660	0.2366	0.4864
50	1.1210	0.2346	0.4844
100	1.2737	0.2336	0.4834
200	1.4253	0.2332	0.4829

Table 5: Parameters for $f_{\|\mathcal{Q}_1^n\|}(x)$

n	mean	var	stdd
25	0.2671	0.2366	0.4864
50	0.2715	0.2346	0.4844
100	0.2737	0.2336	0.4834
200	0.2748	0.2332	0.4829

Table 6: Parameters for $f_{\|\mathcal{Q}_1^n\|}(x)$, shifted

n	mean	var	stdd
25	0.8199	0.0795	0.2819
50	0.9727	0.0785	0.2802
100	1.1243	0.0780	0.2794
200	1.2753	0.0778	0.2789

Table 7: Parameters for $f_{\|\mathcal{Q}_1^n\|}(x)$

n	mean	var	stdd
25	0.1210	0.0795	0.2819
50	0.1232	0.0785	0.2802
100	0.1243	0.0780	0.2794
200	0.1248	0.0778	0.2789

Table 8: Parameters for $f_{\|\mathcal{Q}_1^n\|}(x)$, shifted

5 Concluding remarks.

There have been written numerous papers on the statistical behaviour of norms and condition numbers of random matrices ([1] [2], [3], [7]). However, norms and condition numbers are *worst case indicators* in the error analysis of linear systems. Therefore these estimates are often very pessimistic.
The distributions derived in this paper have a direct application in [5], the convergence analysis of the Krylov solver IDR(s). In the IDR(s) method, real or complex random 'shadow vectors' are used in a similar way as the test-vectors in a Galerkin approximation of a linear system. In [5], the IDR(s) method is compared to the rather expensive full-GMRES procedure, of which a sharp convergence analysis is available. The results obtained in this report provide a good explanation of the observed convergence behaviour of IDR(s).

In the future it may be interesting to study the IDR(s) behaviour for smaller values of s, say $s < 8$. In the corresponding analysis, completely random linear systems are encountered in which the stochastic independency requirements have partly been dropped. However, we have not yet formulated a precise description of these systems.

References

[1] Zizhong Chen and Jack J. Dongarra: *Condition numbers of Gaussian Random matrices;* SIAM Journal on Matrix Analysis and Applications **27(3)**: pp. 603-620, (2005)

[2] Alan Edelman: *Eigenvalues and Condition numbers of Random matrices;* SIAM Journal on Matrix Analysis and Applications **9(4)**: pp. 543-560, (1988)

[3] Malay Ghosh and Bimal K. Sinha: *A simple Derivation of the Wishart Distribution;* The American Statistician, May 2002 **56(2)**: pp. 100-101, (2002)

[4] Martin B. van Gijzen and Peter Sonneveld: *An elegant IDR(s) variant that efficiently exploits bi-orthogonality properties;* Delft University of Technology, Reports of the Department of Applied Mathematical Analysis **08-21**: (2008)

[5] Peter Sonneveld: *On the convergence behaviour of IDR(s);* Delft University of Technology, Reports of the Department of Applied Mathematical Analysis **10-08**: (2010)

[6] Peter Sonneveld and Martin B. van Gijzen: *IDR(s): a family of simple and fast algorithms for solving large nonsymmetric systems of linear equations ;* SIAM J. Sci. and Statist. Comput. **31:2**: pp. 1035-1062, (2008)

[7] Stanislaw J. Szarek: *Condition numbers of random matrices;* Journal of Complexity **7(2)**: pp. 131-149, (1991)

[8] E.T. Whittaker and G.N. Watson: *A Course of Modern Analysis;* Cambridge at the University Press pp. 235-259 (1902-1962)