Mathematical Programming 35 (1986) 140-172
North-Holland

IMPROVED ASYMPTOTIC ANALYSIS OF THE AVERAGE NUMBER OF STEPS PERFORMED BY THE SELF-DUAL SIMPLEX ALGORITHM

Nimrod MEGIDDO

The IBM Almaden Research Center, 650 Harry Road, San José, CA 95120, USA and Department of Statistics, Tel Aviv University, Tel Aviv, Israel

Received 28 November 1983
Revised manuscript received 13 May 1985

In this paper we analyze the average number of steps performed by the self-dual simplex algorithm for linear programming, under the probabilistic model of spherical symmetry. The model was proposed by Smale. Consider a problem of n variables with m constraints. Smale established that for every number of constraints m, there is a constant $c(m)$ such that the number of pivot steps of the self-dual algorithm, $\rho(m, n)$, is less than $c(m)(\ln n)^{m(m+1)}$. We improve upon this estimate by showing that $\rho(m, n)$ is bounded by a function of m only. The symmetry of the function in m and n implies that $\rho(m, n)$ is in fact bounded by a function of the smaller of m and n.

Key words: Probabilistic analysis, self-dual simplex, spherical symmetry.

1. Introduction

In this paper we analyze the average number of pivot steps performed by the self-dual simplex algorithm [12] (also referred to as 'Lemke's algorithm' [15]). The probabilistic analysis of this algorithm was initiated by S. Smale [20, 21]. The probabilistic model, proposed by Smale, is that of complete spherical symmetry. An *equivalent* model for analyzing the self-dual algorithm is one in which all the coefficients of the linear programming problem are sampled independently from the standard normal distribution. The equivalence stems from the fact the number of steps is independent of the radial part of the distribution.

Our analysis in this paper is carried out precisely under Smale's model. We note however that his proof works for a more general model. Smale proved that the expected number of steps $\rho(m, n)$, performed by the self-dual algorithm on a problem of order $m \times n$, satisfies

$$\rho(m, n) \leq c(m) \ln n^{m(m+1)},$$

where $c(m)$ is a constant depending on m. Dantzig [12] conjectured that the expected number of steps of the simplex algorithm (denote it by $\rho^*(m, n)$) satisfies $\rho^*(m, n) \leq c(m)n$ and hence Smale's result can be appraised as proving something close to

Parts of this research were done while the author was visiting Stanford University, XEROX- PARC, Carnegie-Mellon University and Northwestern University and was supported in part by the National Science Foundation under Grants MCS-8300984, ECS-8218181 and ECS-8121741.

Dantzig's conjecture, with reservations concerning the probabilistic model and the algorithm.

Blair [7] proves that the expected number of undominated columns in a problem of order $m \times n$, under an even more general model, is less than $c(m)(\ln n)^{m(m+1)\ln(m+1)+m}$. In general, estimations of numbers of vertices, numbers of undominated columns, or numbers of nonredundant constraints, lead to exponential estimates on the number of steps. Blair's bound is somewhat close to Smale's but naturally cannot produce the result of the present paper since the expected number of undominated columns does tend to infinity.

Another important analysis was carried out by Borgwardt [9, 10, 11]. He worked with a different probabilistic model under which the problem was always feasible. Furthermore, a special algorithm had to be designed in order for the probabilistic analysis to be valid, and this algorithm solved only problems drawn from the particular distribution. Denote the expected number of steps under Borgwardt's model by $\rho^B(m, n)$. We first note that ρ^B is not symmetric in m and n while ρ is. Borgwardt proved that $\rho^B(m, n)$ was polynomial as a function of two variables. Specifically,

$$\rho^B(m, n) \leq cmn^2(n+1)^2,$$

where $c = [(e\pi)/4](\pi/2 + 1/e)$. Thus, Borgwardt's analysis does not show that the number of steps tends to a finite limit when either of the dimensions tends to infinity while the other is fixed.

The work of Smale and Borgwardt (and the article in *Science* magazine [14]) encouraged a number of other researchers to improve analysis and extend it to other algorithms and probabilistic models. The papers of Adler [1] and Haimovich [13] are among those important developments. Since the first version of the present paper different variants of the self-dual algorithm have been observed to require only $O(m^2)$ pivot steps on the average [3, 4, 5, 6, 22]. The author of the present paper has however shown [18] that all these variants can be considered as special cases of the self-dual algorithm. The result in [2] has a spirit similar to that of the present paper. It should be noticed that the result of the present paper does not fall within the framework of [2].

The reader may wish to refer to Smale's papers for the main results leading to the analysis described in the present paper. The background given in Section 2 is rather brief.

2. Preliminaries

Smale [20] presented a formula for $\rho(m, n)$ which was based on probabilities of random rays lying in certain random cones. We will not discuss here the derivation of the main formula. All we say at this point is that we will be estimating probabilities of events related to the formula. First, we introduce the notion of the 'Gaussian volume' (or the 'spherical measure') of a matrix, which arises naturally in Smale's analysis. The Gaussian volume $GV(M)$ of a matrix $M \in R^{m\times n}$ is defined as the probability that a random vector $v \in R^n$, drawn from the standard n-normal distribution, belongs to the convex cone spanned by the columns of M. In general, the Gaussian volume of any measurable set is the probability that v belongs to the set. Our main interest will be evaluating 'expected' volumes of matrices, some of whose entries will be random normal variates, that is, each has the density function

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-(1/2)x^2}.$$

Thus,

$$\begin{aligned} GV(M) &= \frac{1}{(2\pi)^{n/2}} \int_{M^{-1}x \in R^n_+} \exp\{-\tfrac{1}{2}\|x\|^2\} \, dx \\ &= \frac{|\det M|}{(2\pi)^{n/2}} \int_{u \in R^n_+} \exp\{-\tfrac{1}{2}\|Mu\|^2\} \, du. \end{aligned}$$

Using conventional notation, let

$$\Phi(x) = \int_{-\infty}^{x} \phi(t) \, dt.$$

We will assume throughout that all the random entries of our matrices are standard normal variates (usually independent). The non-random entries will be zeros, ones and negative ones.

Smale [20] showed that $\rho(m, n)$ can be expressed as the sum of volumes of matrices of different types. We first describe the types of matrices which arise in the formula for $\rho(m, n)$. The different types of matrices correspond to different types of bases. We then analyze each type separately, showing that the contribution of each is bounded by a function of m only.

To understand the roles of the different types of matrices, we have to look more closely at the self-dual algorithm. Consider the following linear programming problem:

$$\begin{array}{ll} \text{Maximize} & c^{\mathrm{T}}x \\ \text{subject to} & Ax \leq b, \\ & x \geq 0, \end{array}$$

where $x, c \in R^n$, $A \in R^{m\times n}$ and $b \in R^m$. Let

$$M = \left[\begin{array}{c|c} & A^{\mathrm{T}} \\ \hline -A & \end{array}\right]$$

and $q=(c,-b)^{\mathrm{T}}$. A 'primal basis' is a set of m linearly independent m-vectors $p^1,\ldots,p^m$, where each p_i is either a column of $-A$, or a unit vector (consisting of $m-1$ zeros and a single unity). Let $P^i=(0_n,p^i)^{\mathrm{T}}$ (where 0_n is a zero n-vector). Analogously, a 'dual basis' is a set of n linearly independent n-vectors $q^1,\ldots,q^n$ where each q^i is either a column of A^{T} or a unit vector. Denote $Q^j=(q^j,0_m)^{\mathrm{T}}$. Now, the set $B=\{P^1,\ldots,P^m,Q^1,\ldots,Q^n\}$ is called a 'complementary basis' if for every i, $i=1,\ldots,m+n$, either the ith column of m, or the ith unit vector $e^i\in R^{m+n}$ belongs to B. Note that B is linearly independent if we start from primal and dual bases. It follows that in order for B to be a complementary basis, it is necessary that the number of p^i's selected from $-A$ equals the number of q^j's selected from A^{T}. Assuming $m\leqslant n$, this number cannot exceed m. A complementary basis is 'feasible' if

$$B^{-1}q\geqslant 0,$$

in which case $B^{-1}q$ is a vector whose first n components constitute a dual-optimal solution, and the last m components constitute a primal-optimal solution. In the special case where $q=e=(1,\ldots,1)^{\mathrm{T}}\in R^{m+n}$, the columns of the identity matrix I_{m+n} constitute a complementary basis. In general, the self-dual algorithm attempts to find complementary bases relative to points of the form $v(t)=(1-t)e+tq$, t varying from 0 to 1. An equivalent interpretation can be given in terms of artificial bases. A set B^* of $m+n$ linearly independent $(m+n)$-vectors is called an 'artificial basis' if it can be obtained from a complementary basis by replacing one of its columns by $-e$, or if it is itself a complementary basis. Two artificial bases are 'adjacent' if their intersection consists of $m+n-1$ columns. The self-dual algorithm generates a sequence of (feasible) artificial bases, in which every two consecutive ones are adjacent. If the linear programming problem is feasible and bounded then the algorithm terminates with a feasible complementary basis. Otherwise, it recognizes that no such basis exists.

The different types of matrices described below correspond to the different types of artificial bases, depending on the column replaced by $-e$. For example, the quantity $V_1(m,n,k)$ defined below is equal to the probability that a specific artificial basis will be reached by the algorithm; such a basis is characterized by the property that it is obtained from a complementary basis, containing k columns from each of the matrices $-A$ and A^{T}, by replacing one of the unit columns (corresponding to a dual slack) by $-e$. It is easy to verify that the number of different bases of this type in a problem of order $m\times n$ is equal to $(n-k)\binom{n}{k}\binom{m}{k}$. Similarly, $V_2(m,n,k)$ represents the same probability for an artificial basis obtained by replacing one of the units columns (corresponding to a primal slack) by $-e$. The number of different bases of this type is equal to $(m-k)\binom{n}{k}\binom{m}{k}$. Now, $V_3(m,n,k)$ represents the same probability for an artificial basis obtained by replacing a column of A^{T} (expanded with zeros) by $-e$, while $V_4(m,n,k)$ represents bases where a column of $-A$ was replaced.

As a matter of fact, there are two types of matrices that need to be studied. We describe four types, which constitute two equivalent pairs if we interchange the roles of m and n. However, since we analyze the asymptotic behavior when n tends to infinity and m is fixed, such a symmetry does not suffice. Thus, we will estimate expected volumes of matrices of the following types:

The first type of matrices is of the form

$$M_1=\left[\begin{array}{c|c|c|c}
 & -1 & Z & \\ \cline{3-3}
I_{m+n-2k-1} & & & W \\ \hline
 & \vdots & y^{\mathrm{T}} & \\ \cline{3-3}
 & & X & \\ \cline{3-4}
 & -1 & & -X^{\mathrm{T}}
\end{array}\right],$$

where $X\in R^{k\times k}$, $Z\in R^{(n-k-1)\times k}$, $W\in R^{(m-k)\times k}$ and $y\in R^k$ $(0\leqslant k\leqslant m\leqslant n)$. We denote $E(GV(M_1))=V_1(m,n,k)$. A matrix of the type of M_1 arises in the formula for $\rho(m,n)$ a number of times which equals $(n-k)\binom{n}{k}\binom{m}{k}$ and thus we will be interested in the quantity

$$(n-k)\tbinom{n}{k}\tbinom{m}{k}V_1(m,n,k).$$

The second type of matrices is of the form

$$M_2=\left[\begin{array}{c|c|c|c}
 & Z & -1 & \\ \cline{2-2}\cline{4-4}
I_{m+n-2k-1} & & & W \\ \hline
 & X & \vdots & \\ \cline{2-2}\cline{4-4}
 & & & -X^{\mathrm{T}} \\ \cline{4-4}
 & & -1 & y^{\mathrm{T}}
\end{array}\right],$$

where $X\in R^{k\times k}$, $Z\in R^{(n-k)\times k}$, $W\in R^{(m-k-1)\times k}$ and $y\in R^k$ $(0\leqslant k\leqslant m\leqslant n)$. We denote $E(GV(M_2))=V_2(m,n,k)$. Note that $V_2(m,n,k)=V_1(n,m,k)=V_1(n,m,k)$. A matrix of the type of M_2 arises in the formula for $\rho(m,n)$ a number of times which equals $(m-k)\binom{n}{k}\binom{m}{k}$ and thus we will be interested in the quantity

$$(m-k)\tbinom{n}{k}\tbinom{m}{k}V_2(m,n,k).$$

The third type of matrices is of the form

$$M_3 = \left[\begin{array}{c|c|c|c}
 & -1 & Z & \\
I_{m+n-2k} & \vdots & & W \\
\hline
 & & X & \\
 & \cdot & & y^{\mathrm{T}} \\
 & -1 & & -X^{\mathrm{T}}
\end{array}\right],$$

where $X \in R^{k\times(k-1)}$, $Z \in R^{(n-k)\times(k-1)}$, $W \in R^{(m-k)\times k}$ and $y \in R^k$. We denote $E(GV(M_3)) = V_3(m, n, k)$ and will be interested in the quantity

$$n\binom{n-1}{k-1}\binom{m}{k} V_3(m, n, k).$$

since this represents the contribution of $V_3(m, n, k)$ to $\rho(m, n)$.

The fourth type of matrices is of the form

$$M_4 = \left[\begin{array}{c|c|c|c}
 & Z & -1 & \\
I_{m+n-2k} & & \vdots & W \\
\hline
 & y^{\mathrm{T}} & & \\
 & X & & \\
 & & -1 & -X^{\mathrm{T}}
\end{array}\right],$$

where $X \in R^{(k-1)\times k}$, $Z \in R^{(n-k)\times k}$, $W \in R^{(m-k)\times(k-1)}$ and $y \in R^k$. Denote $E(GV(M_4)) = V_4(m, n, k)$. Obviously, $V_4(m, n, k) = V_3(n, m, k)$. We are interested in the quantity

$$m\binom{m-1}{k-1}\binom{n}{k} V_4(m, n, k).$$

The asymptotic behavior when m is fixed and n tends to infinity turns out to be different for V_3 and V_4.

Since $\rho(m, n)$ is symmetric in m and n, we assume that $m \leq n$. It has been shown by Smale that

$$\begin{aligned}\rho(m, n) = \sum_{k=0}^{m} \Bigg(& (n-k)\binom{b}{k}\binom{m}{k} V_1(m, n, k) \\ & + (m-k)\binom{n}{k}\binom{m}{k} V_2(m, n, k) + n\binom{n-1}{k-1}\binom{m}{k} V_3(m, n, k) \\ & + m\binom{m-1}{k-1}\binom{n}{k} V_4(m, n, k)\Bigg).\end{aligned}$$

We use asterisks to denote the square submatrix in the lower-right corner of the matrices M_i $(i = 1, 2, 3)$, for example,

$$M_1^* = \left[\begin{array}{c|c|c} -1 & y^{\mathrm{T}} & \\ \cline{2-2} \vdots & X & \\ \cline{2-3} -1 & & -X^{\mathrm{T}} \end{array}\right],$$

$M_1^* \in R^{(2k+1)\times(2k+1)}$. The following proposition enables us to reduce the dimensions of our integrals from the order of $(m+n)\times(m+n)$-matrices to the order of $(2k)\times(2k)$- and $(2k+1)\times(2k+1)$-matrices.

Proposition 2.1

$$\int_{R^l} \Phi(\alpha_0 - \alpha^{\mathrm{T}} x)\phi(x_1)\cdots\phi(x_l)\,\mathrm{d}x_1\cdots\mathrm{d}x_l = \Phi\left(\frac{\alpha_0}{\sqrt{1+\|\alpha\|^2}}\right)$$

where α and x are in R^l.

Proof. The proof follows by a geometric argument: the integral reflects the probability that a random point $(x_1, \ldots, x_l, y)^{\mathrm{T}}$ satisfies

$$y \leq \alpha_0 - \alpha^{\mathrm{T}} x$$

or, equivalently,

$$\alpha^{\mathrm{T}} x + y \leq \alpha_0.$$

The distance between the hyperplane

$$\{\alpha^{\mathrm{T}} x + y = \alpha_0\}$$

and the origin is equal to

$$\frac{\alpha_0}{\sqrt{1+\|\alpha\|^2}}.$$

The standard multinormal density function is spherically symmetric and the probability that a random point will fall on that side of a hyperplane which contains the origin, is equal to $\Phi(d)$, where d is the distance between the origin and the hyperplane. This implies our claim.

3. An estimation of $V_1(m, n, k)$

We now turn to estimating $V_1(m, n, k)$. Applying Proposition 2.1, let us denote

$$\Delta = \Delta(\lambda, \alpha, \beta, X, y) = \frac{|\det M_1^*|}{(2\pi)^{(k^2+3k+1)/2}} \exp\left\{ -\tfrac{1}{2}\left(\left\| M_1^* \begin{bmatrix} \lambda \\ \alpha \\ \beta \end{bmatrix} \right\|^2 + \|X\|^2 + \|y\|^2 \right) \right\},$$

where $\lambda \in R_+^1$, $\alpha, \beta \in R_+^k$ and $\|X\|^2 = \sum x_{ij}^2$. We sometimes use the abbreviation $|X|$ for $|\det X|$.

Proposition 3.1

$$V_1(m, n, k) = \int \left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}} \right) \right)^{n-k-1} \left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\beta\|^2}} \right) \right)^{m-k} \Delta d(\lambda, \alpha, \beta, X, y),$$

where the integration is over $\lambda, \alpha, \beta \geq 0$, $X \in R^{k\times k}$ *and* $y \in R^k$.

Proof. We are interested in the probability that a random $v \in R^{m+n}$ belongs to the cone spanned by the columns of M_1. Now, under our model M_1 is non-singular with probability one. Assuming this indeed is the case, let $u = M_1^{-1}v$ and represent u as $u = (w^1, w^2, \lambda, \alpha, \beta)^{\mathrm{T}}$ where $w^1 \in R^{n-k-1}$, $w^2 \in R^{m-k}$, $\lambda \in R^1$, $\alpha \in R^k$, $\beta \in R^k$ and $v = (v^1, v^2, v^3)$ where $v^1 \in R^{n-k-1}$, $v^2 \in R^{m-k}$ and $v^3 \in R^{2k+1}$. It follows that $(\lambda, \alpha, \beta)^{\mathrm{T}} = (M_1^*)^{-1}v^3$, $w^1 = v^1 - Z\alpha + \lambda$ and $w^2 = v^2 - W\beta + \lambda$.

By definition,

$$V_1(m, n, k) = \frac{1}{(2\pi)^{(m+n)(k+1)/2}} \int \exp\{-\tfrac{1}{2}(\|Z\|^2 + \|W\|^2 + \|X\|^2 + \|y\|^2 + \|v\|^2)\},$$

where the integration is over all Z, W, X and y (see the dimensions of their respective spaces above) and over v such that $M_1^{-1}v \geq 0$. Substituting $v^3 = M_1^*(\lambda, \alpha, \beta)^{\mathrm{T}}$, we can simplify the integral as follows. When λ, α, β, Z and W are given then $w^1 = v^1 - Z\alpha + \lambda$ and

$$\Pr\{w^1 \geq 0\} = \prod_{i=1}^{n-k-1} \Phi(\lambda - Z^i\alpha)$$

where Z^i is the ith row of Z, and

$$\Pr\{w^2 \geq 0\} = \prod_{i=1}^{m-k} \Phi(\lambda - W^i\beta).$$

Integration over Z and W (using Proposition 1) yields

$$\Pr\{w^1 \geqslant 0\} = \left(\Phi\left(\frac{1}{\sqrt{1+\|\alpha\|^2}}\right)\right)^{n-k-1}$$

(given λ, α and β) and

$$\Pr\{w^2 \geqslant 0\} = \left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\beta\|^2}}\right)\right)^{m-k}.$$

The rest of the proof follows easily.

We now develop a different expression for Δ. First, denote $X^0 = y$ and let $(X^i)^{\mathrm{T}}$ $(i = 1, \ldots, k)$ denote the ith row of the matrix X. Also let

$$S(y, x) = \left|\det \begin{bmatrix} -1 & y^{\mathrm{T}} \\ \vdots & \\ -1 & X \end{bmatrix}\right|, \qquad \mu = \frac{\lambda\alpha}{1+\|\alpha\|^2} \in R^k, \qquad R = (\alpha\alpha^{\mathrm{T}} + I_k)^{-1} \in R^{k\times k},$$

and $1_k = (1, \ldots, 1)^{\mathrm{T}} \in R^k$.

Proposition 3.2

$$\Delta = \frac{S(y, X)|\det X|}{(2\pi)^{(k^2+3k+1)/2}} \exp\left\{-\tfrac{1}{2}\left(\sum_{i=0}^{k} (X^i - \mu)^{\mathrm{T}} R^{-1} (X^i - \mu) + \frac{(k+1)\lambda^2}{1+\|\alpha\|^2} + \|X^{\mathrm{T}}\beta + \lambda \mathbf{1}_k\|^2\right)\right\}.$$

Proof

$$\begin{aligned}
&\|M_1^*(\lambda, \alpha, \beta)^{\mathrm{T}}\|^2 + \|X\|^2 + \|y\|^2 \\
&= (-\lambda + y^{\mathrm{T}}\alpha)^2 + \|-\lambda \mathbf{1}_k + X\alpha\|^2 + \|-\lambda \mathbf{1}_k - X^{\mathrm{T}}\beta\|^2 + \|X\|^2 + \|y\|^2 \\
&= \sum_{i=0}^{k} \left[\left(X^i - \frac{\lambda\alpha}{1+\|\alpha\|^2}\right)^{\mathrm{T}} (\alpha\alpha^{\mathrm{T}} + I)\left(X^i - \frac{\lambda\alpha}{1+\|\alpha\|^2}\right)\right] + \frac{(k+1)\lambda^2}{1+\|\alpha\|^2} \\
&\quad + \|X^{\mathrm{T}}\beta + \lambda \mathbf{1}_k\|^2.
\end{aligned}$$

The rest follows easily. We can now offer interpretations for different quantities (that arise later in our analysis) as follows.

1. $S(y, X)$. As is well known, this is equal to $k!$ times the regular volume in R^k of a simplex whose vertices are $y = X^0, X^1, \ldots, X^k$.

2. $\psi(y, X)$. This is defined by

$$\psi(y, X) = \psi_{\lambda,\alpha}(y, X) = \frac{|\det R|^{-(k+1)/2}}{(2\pi)^{k(k+1)/2}} \exp\left\{-\tfrac{1}{2} \sum_{i=0}^{k} (X^i - \mu)^{\mathrm{T}} R^{-1} (X^i - \mu)\right\},$$

which is equal to the joint probability density function of the variates $X^0, X^1, \ldots, X^k$, assuming these vectors are drawn independently from the k-normal distribution, with mean vector μ and variance-covariance matrix R. Such an assumption turns out to be convenient for our analysis later. We emphasize here that our probabilistic model remains unchanged.

3. $C_\lambda(X)$. This is defined by

$$C_\lambda(X)=\frac{|\det X|}{(2\pi)^{k/2}}\int_{\beta\in R_+^k}\exp\{-\tfrac{1}{2}\|X^{\mathrm{T}}\beta+\lambda\mathbf{1}_k\|^2\}\,\mathrm{d}\beta.$$

The integral reflects the probability that a Gaussian vector $v^*\in R^k$ is representable as $v^*=X^{\mathrm{T}}\beta+\lambda 1_k$ with $\beta\geqslant 0$ (when λ and X are given). In other words, $C_\lambda(X)$ is the Gaussian volume of a cone obtained as follows. Take the cone spanned by the *rows of* X and translate it so that its vertex maps to the point $\lambda 1_k$.

4. $h(\lambda,\alpha)$. This is defined by

$$h(\lambda,\alpha)=\int_{X\in R^{k\times k}}\int_{y\in R^k}S(y,X)C_\lambda(X)\psi_{\lambda,\alpha}(y,X)\,\mathrm{d}y\,\mathrm{d}X$$

which reflects the expectation of a product of two volumes. It is ($k!$ times) the volume of the simplex (whose vertices are $X^0, X^1,\ldots,X^k$) times the Gaussian volume of the cone spanned by the rows of X and then translated as explained above. The expectation is relative to a k-normal distribution $\mathcal{N}(\mu,R)$ from which $X^0, X^1,\ldots,X^k$ are sampled independently.

For an asymptotic analysis of $V_1(m,n,k)$, when n tends to infinity while m and k are fixed, we may look at the following quantity:

$$V_1'(n,k)=\int\left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}}\right)\right)^{n-k-1}\Delta\,\mathrm{d}(\lambda,\alpha,\beta,X,y)=\frac{1}{(2\pi)^{(k+1)/2}}$$
$$\int\left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}}\right)\right)^{n-k-1}\exp\left\{-\frac{1}{2}\frac{(k+1)\lambda^2}{1+\|\alpha\|^2}\right\}|\det R|^{(k+1)/2}h(\lambda,\alpha)\,\mathrm{d}\lambda\,\mathrm{d}\alpha.$$

Note that $|\det R|^{(k+1)/2}$ is inserted here in order to cancel $|\det R|^{-(k+1)/2}$ in the definition of $\psi(y,X)$.

Proposition 3.3

$$\frac{1}{2^{m-k}}V_1'(n,k)\leqslant V_1(m,n,k)\leqslant V_1'(n,k).$$

Proof. This is immediate from Proposition 3.1.

Proposition 3.4
$V_1'(n,k)$

$$=\frac{1}{(2\pi)^{(k+1)/2}}\int_{\delta\in R_+}\int_{\alpha\in R_+^k}(\Phi(\delta))^{n-k-1}\,\mathrm{e}^{-(1/2)(k+1)\delta^2}\frac{h(\delta\sqrt{1+\|\alpha\|^2},\alpha)}{(1+\|\alpha\|^2)^{k/2}}\,\mathrm{d}\delta\,\mathrm{d}\alpha.$$

Proof. The proof goes by substituting $\lambda = \delta\sqrt{1+\|\alpha\|^2}$ (into the definition of $V_1'(n, k)$) observing that

$$|\det R|^{(k+1)/2} = (\det(\alpha\alpha^{\mathrm{T}} + I_k))^{-(k+1)/2} = (1+\|\alpha\|^2)^{-(k+1)/2}.$$

For simplicity of notation let

$$H(\delta,\alpha) = h(\delta\sqrt{1+\|\alpha\|^2}, \alpha)$$

and also let

$$g(\delta) = \int_{\alpha\in R_+^k} \frac{H(\delta, \alpha)}{(1+\|\alpha\|^2)^{k/2}} \, \mathrm{d}\alpha.$$

Thus,

$$V_1'(n, k) = \frac{1}{(2\pi)^{(k+1)/2}} \int_0^\infty (\Phi(\delta))^{n-k-1} e^{-(1/2)(k+1)\delta^2} g(\delta) \, \mathrm{d}\delta.$$

Our goal now will be to show that

$$g(\delta) = \mathrm{O}\left(\frac{1}{\delta^k}\right).$$

To that end, we first estimate $H(\delta, \alpha)$. Recall that $h(\delta, \alpha)$ was interpreted as an expectation of a product of two volumes:

$$h(\delta, \alpha) = \mathscr{E}[S(y, X) C_\lambda(X)]$$

where y as well as the rows of X are sampled independently from the k-normal distribution with mean vector

$$\mu = \frac{\lambda\alpha}{1+\|\alpha\|^2} = \frac{\delta\alpha}{\sqrt{1+\|\alpha\|^2}}$$

and variance-covariance matrix

$$R = (\alpha\alpha^{\mathrm{T}} + I_k)^{-1}.$$

It can be verified that

$$R_{ii} = (\alpha\alpha^{\mathrm{T}} + I_k)_{ii}^{-1} = \frac{1+\|\alpha\|^2 - \alpha_i^2}{1+\|\alpha\|^2}$$

and, for $i \neq j$,

$$R_{ij} = -\frac{\alpha_i\alpha_j}{1+\|\alpha\|^2}.$$

Note that the matrix $\alpha\alpha^{\mathrm{T}} + I_k$ is positive-definite; its eigenvalues are: 1 with multiplicity $k-1$, corresponding to vectors orthogonal to α, and $1+\|\alpha\|^2$ corresponding to the eigenvector α.)

Proposition 3.5. *For any two random variates Y_1 and Y_2.*

$$\mathscr{E}(Y_1 Y_2) \leq \mathscr{E}(Y_1)\mathscr{E}(Y_2) + \sigma(Y_1)\sigma(Y_2).$$

where $\mathscr{E}$ denotes expectation and σ denotes standard deviation.

Proof. This is a well-known fact, usually stated as that the correlation is less than or equal to 1.

Corollary 3.6. $h(\lambda, \alpha) \leq \mathscr{E}[S(y, X)]\mathscr{E}[C_\lambda(X)] + \sigma[S(y, X)]\sigma[C_\lambda(X)]$.

Proposition 3.7. *There exist constants $c = c(k)$ and $d = d(k)$ such that*

$$\mathscr{E}[(S(y, X)] = \frac{c(k)}{\sqrt{1+\|\alpha\|^2}}, \qquad \sigma[(S(y, X)] = \frac{d(k)}{\sqrt{1+\|\alpha\|^2}}.$$

Proof. Recalling that R is positive-definite, let $A \in R^{k\times k}$ be a matrix such that $A^{\mathrm{T}}A = R^{-1}$ ($\det A > 0$) and let

$$A^* = \begin{bmatrix} 1 & 0 \\ 0 & A^T \end{bmatrix} \in R^{(k+1)\times(k+1)}.$$

Obviously, $\det A = \det A^* = \sqrt{1+\|\alpha\|^2}$. Also let $Z^i = A(X^i - \mu)$ ($i = 0, 1, \ldots, k$). Let

$$S(Z) = \left| \det \begin{bmatrix} -1 & Z^0 \\ & Z^1 \\ \vdots & \vdots \\ -1 & Z^k \end{bmatrix} \right|,$$

that is, $S(Z)$ equals $k!$ times the volume of the simplex whose vertices are $Z^0, Z^1, \ldots, Z^k$. The presence of a column of -1's implies that $S(Z) = \det A^* S(y, X)$, so we have

$$\begin{aligned}\mathscr{E}[S(y, X)] &= \frac{|\det R|^{-(k+1)/2}}{(2\pi)^{k(k+1)/2}} \int S(y, X) \exp\left\{-\tfrac{1}{2}\sum_{i=0}^{k} (X^i - \mu)^{\mathrm{T}} R^{-1} (X^i - \mu)\right\} \mathrm{d}y\, \mathrm{d}X \\ &= \frac{|\det R|^{-(k+1)/2}}{(2\pi)^{k(k+1)/2}} \int |\det A^*|^{-1} S(Z) \exp\left\{-\tfrac{1}{2}\sum_{i=0}^{k} \|Z^i\|^2\right\} |\det A^*|^{-(k+1)}\, \mathrm{d}Z.\end{aligned}$$

Letting

$$c(k) = \frac{1}{(2\pi)^{k(k+1)/2}} \int S(Z) \exp\left\{-\tfrac{1}{2}\sum_{i=0}^{k} \|Z^i\|^2\right\} \mathrm{d}Z,$$

that is, $c(k)$ is equal to $k!$ times the expected volume of a k-simplex whose vertices are sampled independently from the *standard* k-normal distribution, we have

$$\mathscr{E}[S(y, X)] = c(k)|\det A^*|^{-1} = \frac{c(k)}{\sqrt{1+\|\alpha\|^2}}.$$

We will later state a precise expression for $c(k)$. The claim about $\sigma[S(y, X)]$ can be proved analogously and a precise expression for $d(k)$ will be derived.

Proposition 3.8. *The expected volume of a simplex whose vertices are sampled independently from the standard multi-normal distribution in R^k is equal to*

$$\frac{1}{k!}2^{k/2}\sqrt{\frac{k+1}{\pi}}\Gamma((k+1)/2).$$

while its variance is equal to

$$\frac{1}{k!^2}\left[(k+1)!-2^k\frac{k+1}{\pi}(\Gamma((k+1)/2))^2\right].$$

Proof. Consider the determinant $S(Z)$ discussed in the proof of Proposition 3.7. We may interpret $S(Z)$ also as the volume in R^{k+1} of a hyperparallelepiped generated by the columns of that matrix. The volume $\mathcal{V}(v^1, \ldots, v^s)$ of a hyperparallelepiped generated by a set of vectors $v^1, \ldots, v^s \in R^l$, relative to the linear subspace that they span, can be described as follows. Consider first the case $s = 1$. Here the volume is obviously $\|v^1\|$. Inductively, consider the linear subspace spanned by $v^1, \ldots, v^{s-1}$ and let v^s be represented as $v^s = v' + v''$ where v' is orthogonal to that subspace while v'' lies in it. Thus, $\mathcal{V}(v^1, \ldots, v^s) = \|v'\|\mathcal{V}(v^1, \ldots, v^{s-1})$. When the vectors $v^1, \ldots, v^s \in R^l$ are drawn independently from the normal distribution, the same formula applies to *expected* volumes. Moreover, given $v^1, \ldots, v^{s-1}$, an equivalent way of sampling v^s is to draw v' and v'' independently from standard normal distributions over their respective subspaces. Thus, v' is generated by sampling $l-s+1$ coordinates independently from standard normal distributions. In other words, the norm of v' is the square root of a χ^2 variable with $l-s+1$ degrees of freedom. It follows [23] that

$$\mathscr{E}[\|v'\|] = \sqrt{2}\frac{\Gamma((l-s+2)/2)}{\Gamma((l-s+1)/2]}.$$

In our case, we start with a fixed vector of norm $\sqrt{k+1}$ and it follows by induction, that the expected volume that we get is

$$\mathscr{E}[S(Z)] = \sqrt{k+1}\sqrt{2}\frac{\Gamma((k+1)/2)}{\Gamma(k/2)}\sqrt{2}\frac{\Gamma(k/2)}{\Gamma((k-1)/2)}\cdots\sqrt{2}\frac{\Gamma(2/2)}{\Gamma(1/2)}$$

$$= \sqrt{k+1}\,2^{k/2}\frac{\Gamma((k+1)/2)}{\Gamma(1/2)}.$$

Our first claim now follows since $\Gamma(0.5) = \sqrt{\pi}$.

A similar analysis applies to the variance. Consider $\mathscr{E}[(S(Z))^2]$. First note that $\mathscr{E}[\|v'\|^2]$ is equal to $l-s+1$. Since we start from a vector whose squared norm is equal to $k+1$, it follows inductively that $\mathscr{E}[S(Z))^2] = (k+1)k\cdots 1 = (k+1)!$. This implies our second claim.

Corollary 3.9

$$\mathscr{E}[S(y,X)]=\sqrt{\frac{2^k(k+1)}{\pi(1+\|\alpha\|^2)}}\,\Gamma((k+1)/2).$$

and

$$\sigma[S(y,X)]=\sqrt{\frac{(k+1)!-((2^k(k+1))/\pi)(\Gamma((k+1)/2))^2}{1+\|\alpha\|^2}}.$$

Before proceeding to the estimation of $C_\lambda(X)$, we find it convenient to substitute

$$\tau=\frac{\alpha}{\sqrt{1+\|\alpha\|^2}}.$$

We have

$$\frac{\partial\tau_i}{\partial\alpha_i}=\frac{1+\|\alpha\|^2-\alpha_i^2}{(1+\|\alpha\|^2)^{3/2}}=(1-\tau_i^2)\sqrt{1-\|\tau\|^2},$$

and, for $i\neq j$,

$$\frac{\partial\tau_i}{\partial\alpha_j}=-\frac{\alpha_i\alpha_j}{(1+\|\alpha\|^2)^{3/2}}=-\tau_i\tau_j\sqrt{1-\|\tau\|^2}.$$

Thus,

$$\left|\det\left[\frac{D\tau}{D\alpha}\right]\right|=(1-\|\tau\|^2)^{k/2}|\det(I_k-\tau\tau^{\mathrm{T}})|=(1-\|\tau\|^2)^{k/2+1}.$$

We now turn to estimating the other volume $C_\lambda(X)$.

Let $K=\{1,\dots,k\}$ and for every $J\subseteq K$ define an event $Ev(J)$ as follows. Recall that the vectors $X^0,\dots,X^k$ are sampled independently from the k-normal distribution with mean vector

$$\mu=\frac{\delta\alpha}{\sqrt{1+\|\alpha\|^2}}$$

and variance-covariance matrix $R=(\alpha\alpha^T+I_k)^{-1}$. Now, $Ev(J)$ will be the event in which J is the set of all the coordinates j such that $X_j^i\geqslant 0$ for all i $(i=0,1,\dots,k)$. In particular, for each $j\notin J$ there exists an i such that $X_j^i<0$. Obviously, the events $Ev(J)$ $(J\subseteq K)$ constitute a partitiion of the sample-space. We now estimate the probability $\Pr(J)$ of the event $Ev(J)$.

Proposition 3.10

$$\Pr(J)\leqslant\left(1-\frac{1}{2^{k+1}}\right)^{k-|J|-1}(k+1)\Phi(-\max_{j\notin J}\{\tau_j\}).$$

Proof. First, observe that each X^i_j is normal with mean

$$\mu_j = \frac{\delta\alpha_j}{\sqrt{1+\|\alpha\|^2}} = \delta\tau_j$$

and variance

$$R_{jj} = (\alpha\alpha^{\mathrm{T}} + I_k)^{-1}_{jj} = \frac{1+\|\alpha\|^2 - \alpha_j^2}{1+\|\alpha\|^2} = 1-\tau_j^2 \leq 1.$$

This can be verified once we know that

$$R_{ij} = -\frac{\alpha_i\alpha_j}{1+\|\alpha\|^2} = \tau_i\tau_j$$

for $i \neq j$. Thus,

$$\Pr(X^i_j < 0) = \Phi\left(\frac{-\mu_j}{\sqrt{R_{jj}}}\right) \leq \Phi(-\delta\tau_j).$$

Now, for every i, the variables X^i_j an X^i_l are negatively correlated (provided $j \neq l$) since

$$\mathrm{Cov}(X^i_j, X^i_l) = R_{jl} = -\tau_j\tau_l.$$

It thus follows that for any fixed i ($i = 0, 1, \ldots, k$) and any set $L \subseteq K$,

$$\Pr\{\forall j \in L)(X^i_j < 0\} \leq \prod_{j\in L} \Phi(-\delta\tau_j).$$

In fact, since the random vectors $X^0, X^1, \ldots, X^k$ are independent, it follows that for any set L of pairs (i, j),

$$\Pr\{(\forall (i,j) \in L)(X^i_j < 0)\} \leq \prod_{(i,j)\in L} \Phi(-\delta\tau_j).$$

Similarly, for every j,

$$\Pr\{(\exists i)(X^i_j < 0)\} \leq 1 - (1-\Phi(-\delta\tau_j))^{k+1} = 1 - (\Phi(\delta\tau_j))^{k+1},$$

so that

$$\begin{aligned}
\Pr(J) &\leq \Pr\{(\forall j \notin J)(\exists i)(X^i_j < 0)\} \leq \prod_{j \notin J} (1 - (\Phi(\delta\tau_j))^{k+1}) \\
&\leq \left(1 - \frac{1}{2^{k+1}}\right)^{k-|J|-1} (1 - (\Phi(\delta \max_{j\notin J}\{\tau_j\}))^{k+1}) \\
&\leq \left(1 - \frac{1}{2^{k+1}}\right)^{k-|J|-1} (k+1)\Phi(-\delta \max_{j\notin J}\{\tau_j\}).
\end{aligned}$$

Corollary 3.11. *For any nonempty set J,*

$$\Pr(J) \leq \left(1 - \frac{1}{2^{k+1}}\right)^{k-|J|},$$

whereas

$$\Pr(\emptyset) \leqslant (k+1)\Phi(-\delta \max_{1\leqslant j\leqslant k}\{\tau_j\}).$$

Proposition 3.12

$$\mathscr{E}[C_\lambda(X)] < (k+1)\Phi(-\delta \max_{1\leqslant j\leqslant k}\{\tau_j\}) + \mathrm{O}(\Phi(-\delta)).$$

Proof. We first estimate the conditional expectation of $C_\lambda(X)$, given the event $\mathrm{Ev}(J)$. Now, given that $X_j^i \geqslant 0$ for every $j \in J$ and i ($i = 0, 1, \ldots, k$), in order for a vector v to be representable as

$$v = X^{\mathrm{T}}\beta + \lambda \mathbf{1}_k = X^{\mathrm{T}}\beta + \delta\sqrt{1+\|\alpha\|^2}\mathbf{1}_k$$

with $\beta \geqslant 0$, it is necessary that

$$v_j \geqslant \delta\sqrt{1+\|\alpha\|^2}$$

for all $j \in J$. If v is a Gaussian vector then the probability of the latter is

$$(\Phi(-\delta\sqrt{1+\|\alpha\|^2}))^{|J|}.$$

Thus,

$$\mathscr{E}[C_\lambda(X) \mid Ev(J)] \leqslant (\Phi(-\delta\sqrt{1+\|\alpha\|^2}))^{|J|} \leqslant (\Phi(-\delta))^{\|J\|}.$$

Finally,

$$\begin{aligned}
\mathscr{E}[C_\lambda(X)] &= \sum_{J\subseteq K} \Pr(J)\mathscr{E}[C_\lambda(X) \mid Ev(J)] \\
&\leqslant \Pr(\emptyset) + \sum_{\emptyset\neq J\subseteq K} \Pr(J)\mathscr{E}[C_\lambda(X) \mid Ev(J)] \\
&\leqslant (k+1)\Phi(-\delta \max_{1\leqslant j\leqslant k}\{\tau_j\}) + \sum_{\emptyset\neq J\subseteq K} \left(1-\frac{1}{2^{k+1}}\right)^{k-|J|} (\Phi(-\delta))^{|J|} \\
&= (k+1)\Phi(-\delta \max_{1\leqslant j\leqslant k}\{\tau_j\}) + \mathrm{O}(\Phi(-\delta)).
\end{aligned}$$

Proposition 3.13

$$\sigma[C_\lambda(X)] < \sqrt{(k+1)\Phi(-\delta \max_{1\leqslant j\leqslant k}\{\tau_j\})} + \mathrm{O}(\Phi(-\delta)).$$

Proof. The proof is essentially the same as that of Proposition 1.12:

$$\begin{aligned}
\sigma^2[C_\lambda(X)] &< \mathscr{E}[(C_\lambda(X))^2] < \Pr(\emptyset) + \sum_{\emptyset\neq J\subseteq K} \Pr(J)\mathscr{E}[(C_\lambda)^2(X) \mid Ev(J)] \\
&\leqslant (k+1)\Phi(-\delta \max_{1\leqslant j\leqslant k}\{\tau_j\}) + \sum_{\emptyset\neq J\subseteq K} \left(1-\frac{1}{k+1}\right)^{k-|J|} (\Phi(-\delta))^{2|J|} \\
&= (k+1)\Phi(-\delta \max_{1\leqslant j\leqslant k}\{\tau_j\}) + \mathrm{O}((\Phi(-\delta))^2).
\end{aligned}$$

This establishes our claim.

Denote

$$P = P_{\delta,\alpha} = \Phi(-\delta \max_{1\le j\le k}\{\tau_j\}).$$

We have established that for some constant $c_1 = c_1(k)$

$$\begin{aligned} H(\delta,\alpha) &\le \mathscr{E}[S(y,X)]\mathscr{E}[C_\lambda(X)] + \sigma[S(y,X)]\sigma[C_\lambda(X)] \\ &\le c_1(k)\frac{\sqrt{P_{\delta,\alpha}}}{\sqrt{1+\|\alpha\|^2}} + \mathrm{O}(\Phi(-\delta). \end{aligned}$$

Proposition 3.14

$$g(\delta) < c_1(k)\int_{\alpha\in R_+^k}(1+\|\alpha\|^2)^{-(k+1)/2}\sqrt{\Phi\left(\frac{-\delta\max_{1\le j\le k}\{\alpha_j\}}{\sqrt{1+\|\alpha\|^2}}\right)}\,\mathrm{d}\alpha + \mathrm{O}(\Phi(-\delta)).$$

Proof. This is a direct consequence of what we have just proved.

Let B_+^k denote the intersection of the k-dimensionl unit ball, centered at the origin, with the non-negative orthant in R^k, that is,

$$B_+^k = \{\tau = (\tau_1,\ldots,\tau_k)\colon \tau_i \ge 0,\ \|\tau\| \le 1\}.$$

Also, let $v(k)$ denote the volume of B_+^k. It follows (see [19]) that

$$v(k) = \frac{\pi^{k/2}}{k2^{k-1}\Gamma(k/2)}.$$

Furthermore, the k-dimensional area of the intersection of the surface of the unit ball with the nonnegative orthant is equal to $kv(k)$.

Proposition 3.15

$$g(\delta) > c_1(k)\int_{\tau\in B_+}\frac{1}{\sqrt{1-\|\tau\|^2}}\sqrt{\Phi(-\delta\max_{1\le j\le k}\{\tau_j\})}\,\mathrm{d}\tau + \mathrm{O}(\Phi(-\delta))$$

Proof. Recall that

$$\left|\det\left[\frac{D\tau}{D\alpha}\right]\right| = (1-\|\tau\|^2)^{k/2+1} \quad\text{and}\quad 1-\|\tau\|^2 = \frac{1}{1+\|\alpha\|^2}.$$

Moreover, the integral

$$\int_{\alpha\in R_+^k}(1+\|\alpha\|^2)^{-(k+1)/2}\,\mathrm{d}\alpha$$

converges. The rest follows easily.

Proposition 3.16. *There exists a constant $C = C(k)$ such that*

$$g(\delta) < C(k)\frac{1}{\delta^k}.$$

Proof. For $\tau \geqslant 0$, we have

$$\|\tau\| \leqslant \sqrt{k} \max_{1 \leqslant j \leqslant k} \tau_j$$

It follows that

$$g(\delta) < c_1(k) \int_{\tau \in B_+^k} \frac{1}{\sqrt{1-\|\tau\|^2}} \sqrt{\Phi\left(-\delta\frac{\|\tau\|}{\sqrt{k}}\right)}\, \mathrm{d}r + \mathrm{O}(\varphi(-\delta)).$$

Since the latter depends only on the norm of τ, it now becomes convenient to switch to polar coordinates. Let $t = \|\tau\|$. Integration over all the angular coordinates yields a constant factor $c_2(k)$ which equals $k/2^k$ times the volume of the unit ball in R^k (or, equivalently, $1/2^k$ of the surface area of that ball). More precisely,

$$c_2(k) = kv(k) = \frac{2\pi^{k/2}}{\Gamma(k/2)}.$$

Thus,

$$g(\delta) < c_1(k)c_2(k) \int_0^1 \sqrt{\frac{\Phi(-\delta t/\sqrt{k})}{1-t^2}}\, t^{k-1}\, \mathrm{d}t + \mathrm{O}(\Phi(-\delta)).$$

Next, we estimate the latter by integrating separately over the intervals $[0, \varepsilon]$ and $[\varepsilon, 1]$, where $0 < \varepsilon < 1$. Thus, we note that the integral

$$\begin{aligned}
\int_0^\varepsilon \sqrt{\frac{\Phi(-\delta t/\sqrt{k})}{1-t^2}}\, t^{k-1}\, \mathrm{d}t &\leqslant \frac{1}{\sqrt{1-\varepsilon^2}} \int_0^\varepsilon \sqrt{\Phi(-\delta t/\sqrt{k})}\, t^{k-1}\, \mathrm{d}t \\
&= \frac{1}{\delta^k\sqrt{1-\varepsilon^2}} \int_0^{\varepsilon\delta} \sqrt{\Phi(-y/\sqrt{k})}\, y^{k-1}\, \mathrm{d}y \\
&< \frac{1}{\delta^k\sqrt{1-\varepsilon^2}} \int_0^{\infty} \sqrt{\Phi(-y/\sqrt{k})}\, y^{k-1}\, \mathrm{d}y.
\end{aligned}$$

We note that the integral

$$\int_0^\infty \sqrt{\Phi(-y/\sqrt{k})}\, y^{k-1}\, \mathrm{d}y$$

converges, since $\Phi(-y/\sqrt{k})$ is asymptotically equal to $\sqrt{k}\varphi(y/\sqrt{k})/y$, as y tends to infinity. We will later give an upper bound for this integral as a function of k. The second integral is of a lower order of magnitude, in view of the following:

$$\int_{\varepsilon}^{1} \sqrt{\frac{\Phi(-\delta t/\sqrt{k})}{1-t^2}}\, t^{k-1}\,\mathrm{d}t \leqslant \sqrt{\Phi(-\delta\varepsilon/\sqrt{k})} \int_{\varepsilon}^{1} \frac{\mathrm{d}t}{\sqrt{1-t^2}}$$

$$= \sqrt{\Phi(-\delta\varepsilon/\sqrt{k})}\left(\frac{\pi}{2} - \sin^{-1}(\varepsilon)\right) = \mathrm{o}\left(\frac{1}{\delta^k}\right).$$

It follows that

$$g(\delta) < c_1(k)c_2(k)c_3(k)\frac{1}{\delta^k} + \mathrm{o}\left(\frac{1}{\delta^k}\right)$$

where

$$c_3(k) = \int_0^{\infty} \sqrt{\Phi(-y/\sqrt{k})}\, y^{k-1}\,\mathrm{d}y.$$

Remark. The constant $c_3(k)$ grows quite rapidly with k. Obviously,

$$c_3(k) = k^{k/2} \int_0^{\infty} \sqrt{\Phi(-x)}\, x^{k-1}\,\mathrm{d}x.$$

The integral grows like $2^k(k/2)!$ On the other hand, it does not seem that a sharper estimate of $\Pr(J)$ (see Proposition 9) can yield a better bound. More specifically, we can use the initial estimate

$$\Pr(J) \leqslant \prod_{j \notin J} (1 - (\Phi(\delta\tau_j))^{k+1}).$$

This leads to

$$g(\delta) < c_1(k) \int_{\tau \in B_+^k} \frac{1}{\sqrt{1-\|\tau\|^2}} \sqrt{\prod_{j=1}^{k} (1 - (\Phi(\delta\tau_j))^{k+1})}\,\mathrm{d}\tau + \mathrm{O}(\Phi(-\delta)).$$

We now integrate separately over $\|\tau\| \leqslant \varepsilon$ and $\|\tau\| \geqslant \varepsilon$, the former being the dominant term. Thus,

$$\int_{\tau \in \varepsilon B_+^k} \frac{1}{\sqrt{1-\|\tau\|^2}} \sqrt{\prod_{j=1}^{k} (1 - (\Phi(\delta\tau_j))^{k+1})}\,\mathrm{d}\tau$$

$$< \frac{1}{\delta^k\sqrt{1-\varepsilon^2}} \int_{\eta \in \varepsilon\delta B_+^k} \sqrt{\prod_{j=1}^{k} (1 - (\Phi(\eta_j))^{k+1})}\,\mathrm{d}\eta,$$

and the equivalent of $c_3(k)$ becomes

$$c_3'(k) = \lim_{\delta\to\infty} \int_{\eta\in\delta B_+^k} \sqrt{\prod_{j=1}^{k} (1-(\Phi(\eta_j))^{j+1})}\, \mathrm{d}\eta$$
$$= \int_{\eta\in R_+^k} \sqrt{\prod_{j=1}^{k} (1-(\Phi(\eta_j))^{k+1})}\, \mathrm{d}\eta = \left(\int_0^\infty \sqrt{1-(\Phi(y))^{k+1}}\, \mathrm{d}y\right)^k.$$

Even the latter grows super-exponentially with k.

We are now ready to show that for every fixed k, the sequence $n^{k+1}V_1(m, n, k)$ $(n = 1, 2, \ldots)$ is bounded.

Theorem 3.17. *For every fixed m and k there exists a finite number $V_1''(m, k)$ such that*

$$n^{k+1}V_1(m, n, k) \leqslant V_1''(m, k).$$

Proof. In view of our last proposition we need to consider the following integral:

$$I(n, k) = n^{k+1} \int_0^\infty (\Phi(\delta))^{n-k-1}\, \mathrm{e}^{-(1/2)(k+1)\delta^2} g(\delta)\, \mathrm{d}\delta.$$

Note that, for any finite D and any fixed k,

$$\lim_{n\to\infty} n^{k+1} \int_0^D (\Phi(\delta))^{n-k-1}\, e^{-(1/2)(k+1)\delta^2} g(\delta)\, \mathrm{d}\delta = 0.$$

This is true because the integrand tends exponentially to zero. Thus, it suffices to consider the following integral:

$$\int_D^\infty (\Phi(\delta))^{n-k-1}\, \mathrm{e}^{-(1/2)(k+1)\delta^2} \frac{1}{\delta^k}\, \mathrm{d}\delta.$$

On the other hand, by Mill's ratio,

$$\frac{\mathrm{e}^{-(1/2)\delta^2}}{\sqrt{2\pi}\,\delta} \sim 1 - \Phi(\delta).$$

More precisely, the following estimates can be obtained by integrating $\varphi(t)$ (from δ to infinity) by parts, using the identity $\varphi(t) = (t\varphi(t))(1/t)$, and then repeating with $\varphi(t)(1/t^2)$ as the new integrand. Thus, for $\delta > 0$,

$$\phi(\delta)\left(\frac{1}{\delta} - \frac{1}{\delta^3}\right) < 1 - \Phi(\delta) < \frac{\phi(\delta)}{\delta}.$$

Also, for $\delta \geqslant D > 1$,

$$\frac{\mathrm{e}^{-(1/2)\delta^2}}{\delta} < \frac{\sqrt{2\pi}D^2}{D^2-1}(1-\Phi(\delta)).$$

We have argued that D can be chosen arbitrarily large without affecting the asymptotic behavior of the integral as n tends to infinity, With increasing values of D we get decreasing upper bounds on limsup $I(n, k)$. The limit of these upper bounds shows that limsup $I(n, k)$ is not greater than

$$\lim_{n\to\infty}(2\pi)^{(k+1)/2}n^{k+1}\int_D^{\infty}(\Phi(\delta))^{n-k-1}(1-\Phi(\delta))^k\phi(\delta)\,\mathrm{d}\delta.$$

Now

$$\int_{-\infty}^{\infty}(\Phi(\delta))^{n-k-1}(1-\Phi(\delta))^k\phi(\delta)\,\mathrm{d}\delta=\int_0^1 x^{n-k-1}(1-x)^k\,\mathrm{d}x=\frac{1}{n\binom{n-1}{k}}.$$

This suffices for proving that our sequence is bounded.

We finally turn to bounding the asymptotic contribution of $V_1(m, n, k)$ to $\rho(m, n)$ when n tends to infinity, while m and k are fixed. Let us denote it by

$$\rho_1(m, n, k)=(m+n-2k)\binom{n}{k}\binom{m}{k}V_1(m, n, k).$$

It can be verified that the constants we have encountered during our analysis were the following: $(2\pi)^{-(k+1)/2}$ (this one cancels out in the final integral), $c_1(k)$, $c_2(k)$, $c_3(k)$ and the estimate of the final integral was

$$\frac{1}{(n+1)\binom{n}{k}}.$$

It follows that the product of all coefficients yields

$$\rho_1(m, m, k)<\binom{m}{k}c_1(k)c_2(k)c_3(k).$$

The product of $c_1(k)c_2(k)c_3(k)$ grows super-exponentially with k. On the other hand, in order for a sum of the form

$$\sum_{k=0}^{m}\binom{m}{k}C(k)$$

to be polynomial in m, it is necessary for the sequence $C(k)$ to tend to zero super-exponentially. Thus, we are at this point quite far from proving that $\rho(m, n)$ tends to a polynomial function of m as n tends to infinity.

4. An estimation of $V_3(m, n, k)$

The estimation of $V_3(m, n, k)$ goes like that of $V_1(m, n, k)$, whereas $V_2(m, n, k)$ and $V_4(m, n, k)$ are somewhat different. Our proofs in the present section are rather concise. For more detail refer to the preceding section. First, we recall that

$$M_3^* = \left[\begin{array}{c|c|c} -1 & X & \\ \vdots & & \\ \cline{2-3} & & y^{\mathrm{T}} \\ \cline{3-3} & & -X^{\mathrm{T}} \\ -1 & & \end{array}\right]$$

where $M_3 \in R^{2k\times 2k}$, $X \in R^{k\times(k=1)}$ and $y \in R^k$. Denote

$$\Delta = \Delta(\lambda, \alpha, \beta, X, y) = \frac{|\det M_3^*|}{(2\pi)^{k(k+2)/2}} \exp\left\{ -\tfrac{1}{2}\left(\left\| M_3^* \begin{bmatrix} \lambda \\ \alpha \\ \beta \end{bmatrix} \right\|^2 + \|X\|^2 + \|y\|^2 \right) \right\},$$

where $\lambda \in R_+^1$, $\alpha \in R_+^{k-1}$, $\beta \in R_+^k$. We know that

$$V_3(m, n, k) = \int \left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}} \right) \right)^{n-k} \left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\beta\|^2}} \right) \right)^{m-k} \Delta \, \mathrm{d}(\lambda, \alpha, \beta, X, y),$$

where the integration is over $\lambda, \alpha, \beta \geq 0$ and all $X \in R^{k\times(k-1)}$ and $y \in R^k$. Denoting the ith row of X by $(X^i)^{\mathrm{T}}$, the exponential factor simplifies as follows.

$$\begin{aligned} &\|M_3^*(\lambda, \alpha, \beta)^{\mathrm{T}}\|^2 + \|X\|^2 + \|y\|^2 = \|-\lambda \mathbf{1}_k + X\alpha\|^2 + (-\lambda + y^{\mathrm{T}}\beta)^2 + \|-\lambda \mathbf{1}_{k-1} - X^{\mathrm{T}}\beta\|^2 \\ &\quad + \|X\|^2 + \|y\|^2 = \sum_{i=1}^{k-1} \left[\left(X^i - \frac{\lambda\alpha}{1+\|\alpha\|^2} \right)^{\mathrm{T}} (\alpha\alpha^{\mathrm{T}} + I) \left(X^i - \frac{\lambda\alpha}{1+\|\alpha\|^2} \right) \right] \\ &\quad + \|y\|^2 + \frac{k\lambda^2}{1+\|\alpha\|^2} + \|X^{\mathrm{T}}\beta + \lambda \mathbf{1}_{k-1}\|^2 + (-\lambda + y^{\mathrm{T}}\beta)^2. \end{aligned}$$

Denote

$$S(X) = \left| \det \begin{bmatrix} -1 & \\ \vdots & X \\ -1 & \end{bmatrix} \right|.$$

Obviously, $S(X)$ equals $(k-1)!$ times the (regular) volume of a simplex whose vertices are the rows of X. Also, denote by X^* a $(k\times k)$-matrix whose last row is y^{T} and the rest of whose rows are those of $-X^{\mathrm{T}}$. Now let

$$C_\lambda(y, X) = \frac{|\det X^*|}{(2\pi)^{k/2}} \int_{\beta \in R_+^k} \exp\{-\tfrac{1}{2}((y^{\mathrm{T}}\beta - \lambda)^2 + \|X^{\mathrm{T}}\beta + \lambda \mathbf{1}_{k-1}\|^2)\} \, \mathrm{d}\beta.$$

It can be seen that $C_\lambda(y, X)$ is the (Gaussian) volume of a cone defined as follows. For each i, $i = 1, \ldots, k$, append to the ith row of X a kth coordinate, $-y_i$, and denote the resulting vector by Y^i. Now translate the cone spanned by the vectors $Y^1, \ldots, Y^k$, so that its vertex maps to the point $\lambda \mathbf{1}_k$. Next, denote

$$\psi(y, X) = \psi_{\lambda,\alpha}(y, X) = \frac{|\det R|^{-k/2}}{(2\pi)^{(1/2)k^2}} \exp\left\{-\tfrac{1}{2}\left(\sum_{i=1}^{k} (X^i - \mu)^{\mathrm{T}} R^{-1}(x^i - \mu) + \|y\|^2\right)\right\},$$

where $R = (\alpha\alpha^{\mathrm{T}} + I_{k-1})^{-1} \in R^{(k-1)\times(k-1)}$. Thus, we may interpret $\psi(y, X)$ as the joint probability density function of vectors $Y^1, \ldots, Y^k$. Assuming these vectors are drawn, independently, from a k-normal distribution with mean vector μ ($\mu_j = \delta\tau_j$ for $j = 1, \ldots, k-1$ and $\mu_k = 1$) and variance-covariance matrix

$$R^* = \begin{bmatrix} R & \\ & 1 \end{bmatrix}.$$

Now, define

$$h(\lambda, \alpha) = \int_{X \in R^{k\times(k-1)}} \int_{y \in R^k} S(X) C_\lambda(y, X) \psi_{\lambda,\alpha}(y, X)\, \mathrm{d}y\, \mathrm{d}X$$

which reflects the expectation of a product of two volumes. Next, define

$$\begin{aligned} V_3'(n, k) &= \int \left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}}\right)\right)^{n-k} \Delta d(\lambda, \alpha, \beta, X, y) \\ &= \frac{1}{(2\pi)^{k/2}} \int \left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}}\right)\right)^{n-k} \exp\left\{-\frac{1}{2}\frac{k\lambda^2}{1+\|a\|^2}\right\} |\det R|^{k/2} h(\lambda, \alpha)\, \mathrm{d}\lambda\, \mathrm{d}\alpha. \end{aligned}$$

We now substitute

$$\lambda = \delta\sqrt{1+\|\alpha\|^2} \quad \text{and} \quad \alpha = \tau\sqrt{1+\|\alpha\|^2} \quad (\tau \in R^{k-1}).$$

This yields

$$V_3'(n, k) = \frac{1}{(2\pi)^{k/2}} \int_{\delta \in R_+} \int_{\tau \in B_+^{k-1}} (\Phi(\delta))^{n-k}\, \mathrm{e}^{-(1/2)k\delta^2} \frac{H(\delta, \tau)}{1 - \|\tau\|^2}\, \mathrm{d}\tau\, \mathrm{d}\delta,$$

where

$$H(\delta, \tau) = h(\delta\sqrt{1+|\alpha\|^2}, \tau\sqrt{1+\|\alpha\|^2}) = h(\lambda, \alpha).$$

We now define

$$g(\delta) = \int_{\tau \in B_+^{k-1}} \frac{H(\delta, \tau)}{1 - \|\tau\|^2}\, \mathrm{d}\tau$$

and our goal is to show that

$$g(\delta) = \mathrm{O}\left(\frac{1}{\delta^{k-1}}\right).$$

We rely on the inequality

$$h(\lambda, \alpha) \leq \mathscr{E}[S(X)]\mathscr{E}[C_\lambda(y, X)] + \sigma[S(X)]\sigma[C_\lambda(y, X)].$$

First, by Propositions 3.7 and 3.8, we know that

$$\mathscr{E}[S(X)] = \sqrt{\frac{2^{k-1}k}{\pi(1+\|\alpha\|^2)}}\Gamma(k/2)$$

and

$$\sigma[S(X)] = \sqrt{\frac{k! - ((2^{k-1}k)/\pi)(\Gamma(k/2))^2}{1+\|\alpha\|^2}}.$$

Secondly, we estimate $\mathscr{E}[C_\lambda(y, X)]$ and $\sigma[C_\lambda(y, X)]$ along the lines suggested in the preceding section. We define $EV(J)$ to be the event in which $J \subseteq K$ is the set of all the coordinates j in which each of the vectors $Y^1, \ldots, Y^k$ has a non-negative component. We note that

$$\Pr(Y_j^i < 0) = \Phi(-\delta\tau_j) \quad (j = 1, \ldots, k-1) \quad \text{and} \quad \Pr(Y_k^i < 0) = \tfrac{1}{2}.$$

It is convenient to define $\tau_k = 0$ and then the estimates from the preceeding section follow through with the k vectors $Y^1, \ldots, Y^k$ replacing the $k+1$ vectors $X^0, X^1, \ldots, X^k$. In particular, for a nonempty J,

$$\Pr(J) \leq \left(1 - \frac{1}{2^k}\right)^{k-|J|},$$

whereas

$$\Pr(\emptyset) \leq k\Phi(-\delta \max_{1\leq j\leq k}\{\tau_j\}).$$

It follows that

$$\mathscr{E}[C_\lambda(y, X)] < k\Phi(-\delta \max_{1\leq j\leq k}\{\tau_j\}) + O(\Phi(-\delta))$$

and

$$\sigma[C_\lambda(y, X)] < \sqrt{k\Phi(-\delta \max_{1\leq j\leq k}\{\tau_j\})} + O(\Phi(-\delta)).$$

Thus,

$$H(\delta, \tau) \leq 2k\sqrt{(k-1)!}\sqrt{(1-\|\tau\|^2)\Phi(-\delta \max_{1\leq j\leq k}\{\tau_j\})} + O(\Phi(-\delta)).$$

We now get

$$g(\delta) < 2k\sqrt{(k-1)!}\int_{\tau\in B_+^{k-1}} \frac{1}{\sqrt{1-\|\tau\|^2}}\sqrt{k\Phi(-\delta \max_{1\leq j\leq k}\{\tau_j\})}\,\mathrm{d}\tau + O(\Phi(-\delta)),$$

from which it follows that

$$g(\delta)=\mathrm{O}\left(\frac{1}{\delta^{k-1}}\right).$$

The rest is essentially the same as in the analysis of $V_1(m, n, k)$.

Theorem 4.1. *For every fixed m and k there exists a finite number $V_3''(m, k)$ such that*

$$n^k V_3(m, n, k) \leqslant V_3''(m, k).$$

Proof. Since $g(\delta)=\mathrm{O}(1/\delta^{k-1})$, the present theorem follows from Theorem 3.17 with $k-1$ replacing k. We note that the bound has the same order of magnitude as that of Theorem 3.17.

5. The estimation of $V_2(m, n, k)$ and $V_4(m, n, k)$

The analysis in this section is similar to those of the previous cases, even though an unexpected complication (with a surprising result) does arise, as we shall see later. We start with $V_4(m, n, k)$. Consider the matrix

$$M_4^* = \left[\begin{array}{c|c|c} y^{\mathrm{T}} & -1 & \\ \cline{1-1} X & \vdots & \\ \cline{1-1}\cline{3-3} & & -X^{\mathrm{T}} \\ & -1 & \end{array}\right],$$

where $M_4^* \in R^{2k\times 2k}$, $X \in R^{(k-1)\times k}$ and $y \in R^k$. Denote

$$\Delta=\Delta(\lambda, \alpha, \beta, X, y)=\frac{|\det M_4^*|}{(2\pi)^{k(k+2)/2}}\exp\left\{-\tfrac{1}{2}\left(\left\|M_4^*\begin{bmatrix}\alpha\\ \lambda\\ \beta\end{bmatrix}\right\|^2+\|X\|^2+\|y\|^2\right)\right\},$$

where $\lambda \in R_+^1$, $\alpha \in R_+^k$, $\beta \in R_+^{k-1}$. We know that

$$V_4(m, n, k)=\int\left(\varphi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}}\right)\right)^{n-k}\left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\beta\|^2}}\right)\right)^{m-k}\Delta\ \mathrm{d}(\lambda, \alpha, \beta, X, y),$$

where the integration is over $\lambda, \alpha, \beta \geqslant 0$ and all $X \in R^{(k-1)\times k}$ and $y \in R^k$. Denote the ith row of X by $(X^i)^{\mathrm{T}}$ and, for the convenience of notation, let $X^k = y$. Now, the exponential factor simplifies as follows,

$$
\begin{aligned}
&\|M_4^*(\alpha, \lambda, \beta)^{\mathrm{T}}\|^2 + \|X\|^2 + \|y\|^2 \\
&= (-\lambda + y^{\mathrm{T}}\alpha)^2 + \|-\lambda \mathbf{1}_{k-1} + X\alpha\|^2 + \|-\lambda \mathbf{1}_k - X^{\mathrm{T}}\beta\|^2 + \|X\|^2 + \|y\|^2 \\
&= \sum_{i=1}^{k} \left[\left(X^i - \frac{\lambda\alpha}{1+\|\alpha\|^2} \right)^{\mathrm{T}} (\alpha\alpha^{\mathrm{T}} + I) \left(X^i - \frac{\lambda\alpha}{1+\|\alpha\|^2} \right) \right] \\
&\quad + \frac{k\lambda^2}{1+\|\alpha\|^2} + \|X^{\mathrm{T}}\beta + \lambda \mathbf{1}_k\|^2.
\end{aligned}
$$

Denote by $S(y, X)$ the absolute value of the determinant of a matrix whose columns are $X^1, \ldots, X^{k-1}, X^k = y$. If the vectors $X^1, \ldots, X^k$ are drawn independently from a k-normal distribution with mean vector μ and variance-covariance matrix $R = (\alpha\alpha^{\mathrm{T}} + I_k)^{-1}$, then it can be shown (as we did in Section 3) that

$$
\mathscr{E}[S(y, X) = \sqrt{\frac{2^k}{\pi(1+\|\alpha\|^2)}}\, \Gamma((k+1)/2)
$$

and

$$
\sigma[S(y, X)] = \sqrt{\frac{k! - (2^k/\pi)(\Gamma((k+1)/2))^2}{1+\|\alpha\|^2}}.
$$

As in the previous sections,

$$
\mu = \frac{\lambda\alpha}{\sqrt{1+\|\alpha\|^2}}.
$$

Next, let

$$
\psi(y, X) = \psi_{\lambda,\alpha}(y, X) = \frac{|\det R|^{-k/2}}{(2\pi)^{(1/2)k^2}} \exp\left\{ -\tfrac{1}{2} \sum_{i=1}^{k} (X^i - \mu)^{\mathrm{T}} R^{-1} (X^i - \mu) \right\},
$$

which is equal to the joint probability density function of the variates $X^1, \ldots, X^k$ assuming these vectors are drawn independently from a k-normal distribution with mean vector μ and variance-covariance matrix R.

Now, let

$$
Q(X) = \begin{bmatrix} -1 & \\ \vdots & X^{\mathrm{T}} \\ -1 & \end{bmatrix},
$$

and define

$$
C_\lambda(X) = \frac{|\det Q(X)|}{(2\pi)^{(k-1)/2}} \int_{\beta \in R_+^{k-1}} \exp\{-\tfrac{1}{2}\|X^{\mathrm{T}}\beta + \lambda \mathbf{1}_k\|^2\}\, \mathrm{d}\beta.
$$

Here we cannot interpret $C_\lambda(X)$ as simply as we did in the previous sections, but we can successfully bound it as follows. First, let x^j denote the jth column of X

for $j=1,\dots,k$. Now, denote by $Q_j(X)$ a matrix of order $(k-1)\times(k-1)$ whose rows are $(x^1)^{\mathrm{T}},\dots,(x^{j-1})^{\mathrm{T}},(x^{j+1})^{\mathrm{T}},\dots,(x^k)^{\mathrm{T}}$. Obviously,

$$|\det Q(X)|\leq|\det Q_1(X)|+\cdots+|\det Q_k(X)|.$$

Moreover,

$$\|X^{\mathrm{T}}\beta+\lambda\mathbf{1}_k\|^2=\sum_{j=1}^{k}((x^j)^{\mathrm{T}}\beta+\lambda)^2.$$

It follows that

$$\begin{aligned}C_\lambda(X)&\leq\frac{\sum_{j=1}^k|\det Q_j(X)|}{(2\pi)^{(k-1)/2}}\int_{\beta\in R_+^{k-1}}\exp\left\{-\tfrac12\sum_{\nu=1}^{k}((x^\nu)^{\mathrm{T}}\beta+\lambda)^2\right\}\mathrm{d}\beta\\&\leq\sum_{j=1}^{k}\frac{|\det Q_j(X)|}{(2\pi)^{(k-1)/2}}\int_{\beta\in R_+^{k-1}}\exp\left\{-\tfrac12\sum_{\substack{\nu=1\\\nu\neq j}}^{k}((x^\nu)^{\mathrm{T}}\beta+\lambda)^2\right\}\mathrm{d}\beta\\&\leq\sum_{j=1}^{k}\frac{|\det Q_j(X)|}{(2\pi)^{(k-1)/2}}\int_{\beta\in R_+^{k-1}}\exp\{-\tfrac12\|Q_j(X)\beta+\lambda\mathbf{1}_{k-1}\|^2\}\,\mathrm{d}\beta.\end{aligned}$$

We can now apply our estimates from Section 3 to the integrals

$$C_\lambda^j(X)=\frac{|\det Q_j(X)|}{(2\pi)^{(k-1)/2}}\int_{\beta\in R_+^{k-1}}\exp\{-\tfrac12\|Q_j(X)\beta+\lambda\mathbf{1}_{k-1}\|^2\}\,\mathrm{d}\beta.$$

After the usual substitution

$$\lambda=\delta\sqrt{1+\|\alpha\|^2}\quad\text{and}\quad\alpha=\tau\sqrt{1+\|\alpha\|^2}\quad(\tau\in R^k),$$

we obtain

$$\mathscr{E}[C_\lambda^j(X)]\leq k\Phi(-\delta\max_{1\leq i\leq k}\{\tau_i\})+\mathrm{O}(\Phi(-\delta))$$

and

$$\sigma[C_\lambda^j(X)]<\sqrt{k\Phi(-\delta\max_{1\leq i\leq k}\{\tau_i\})}+\mathrm{O}(\Phi(-\delta)).$$

Finally, for $C_\lambda(X)$ we get

$$\mathscr{E}[C_\lambda(X)]\leq k^2\Phi(-\delta\max_{1\leq j\leq k}\{\tau_j\})+\mathrm{O}(\Phi(-\delta))$$

and

$$\sigma[C_\lambda(X)]<k\sqrt{k\Phi(-\delta\max_{1\leq j\leq k}\{\tau_j\})}+\mathrm{O}(\Phi(-\delta)).$$

Returning to the main proof, let

$$h(\lambda,\alpha)=\int_{X\in R^{(k-1)\times k}}\int_{y\in R^k}S(y,X)C_\lambda(X)\psi_{\lambda,\alpha}(y,X)\,\mathrm{d}y\,\mathrm{d}X$$

which, as before, reflects the expectation of a product of two volumes. Next, define

$$V_4'(n,k)=\int\left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}}\right)\right)^{n-k}\Delta\,\mathrm{d}(\lambda,\alpha,\beta,X,y)$$

$$=\frac{1}{(2\pi)^{k/2}}\int\left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}}\right)\right)^{n-k}\exp\left\{-\frac{1}{2}\frac{k\lambda^2}{1+\|\alpha\|^2}\right\}$$

$$\times|\det R|^{k/2}h(\lambda,\alpha)\,\mathrm{d}\lambda\,\mathrm{d}\alpha.$$

We know that

$$V_4'(n,k)=\frac{1}{(2\pi)^{k/2}}\int_{\delta\in R_+}\int_{\tau\in B_+^k}(\Phi(\delta))^{n-k}\,\mathrm{e}^{-(1/2)k\delta^2}\frac{H(\delta,\tau)}{1-\|\tau\|^2}\,\mathrm{d}\tau\,\mathrm{d}\delta,$$

where

$$H(\delta,\tau)=h(\delta\sqrt{1+\|\alpha\|^2},\tau\sqrt{1+\|\alpha\|^2})=h(\lambda,\alpha).$$

We now define

$$g(\delta)=\int_{\tau\in B_+^k}\frac{H(\delta,\tau)}{1-\|\tau\|^2}\,\mathrm{d}\tau$$

and for our purposes it would be sufficient to show that

$$g(\delta)=\mathrm{O}\left(\frac{1}{\delta^{k-1}}\right).$$

However, it is easy to see that our previous bounds now lead to

$$g(\delta)=\mathrm{O}\left(\frac{1}{\delta^{k}}\right).$$

Recall that,

$$V_4'(n,k)=\frac{1}{(2\pi)^{(k/2)}}\int_0^\infty(\Phi(\delta))^{n-k}\,\mathrm{e}^{-(1/2)k\delta^2}g(\delta)\,\mathrm{d}\delta.$$

Now, consider the integral:

$$I(n,k)=n^k\int_0^\infty(\Phi(\delta))^{n-k}\,\mathrm{e}^{-(1/2)k\delta^2}g(\delta)\,\mathrm{d}\delta.$$

In analogy to the previous sections, for every $D\geqslant 1$ and k,

$$\lim_{n\to\infty}I(n,k)=\lim_{n\to\infty}n^k\int_D^\infty(\Phi(\delta))^{n-k}\,\mathrm{e}^{-(1/2)k\delta^2}\left(\frac{1}{\delta^k}\right)\mathrm{d}\delta.$$

It follows that

$$\lim_{n\to\infty}I(n,k)=\lim_{n\to\infty}(2\pi)^{k/2}n^k\int_D^\infty(\Phi(\delta))^{n-k}(1-\Phi(\delta))^{k-1}\left(\frac{1}{\delta}\right)\phi(\delta)\,\mathrm{d}\delta.$$

The presence of $1/\delta$ in this formula implies that $I(n, k)$ tends to *zero* as n tends to infinity. Thus, we have the following theorem.

Theorem 5.1. *For every fixed m and k,*

$$\lim_{n\to\infty} n^k V_4(m, n, k) = 0.$$

The analysis of $V_2(m, n, k)$ is similar. Consider the matrix

$$M_2^* = \left[\begin{array}{c|c|c} X & -1 & \\ \cline{1-1}\cline{3-3} & \vdots & -X^{\mathrm{T}} \\ \cline{3-3} & & y^{\mathrm{T}} \\ & -1 & \end{array}\right]$$

($M_2^* \in R^{(2k+1)\times(2k+1)}$). Denote

$$\Delta = \Delta(\lambda, \alpha, \beta, X, y) = \frac{|\det M_2^*|}{(2\pi)^{(k^2+3k+1)/2}} \exp\left\{-\tfrac{1}{2}\left(\left\| M_2^* \begin{bmatrix}\alpha\\ \lambda\\ \beta\end{bmatrix}\right\|^2 + \|X\|^2 + \|y\|^2\right)\right\},$$

where $\lambda \in R_+^1$, $\alpha \in R_+^k$, $\beta \in R_+^k$. We know that

$$V_2(m, n, k) = \int \left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\alpha\|^2}}\right)\right)^{n-k} \left(\Phi\left(\frac{\lambda}{\sqrt{1+\|\beta\|^2}}\right)\right)^{m-k-1} \Delta \, \mathrm{d}(\lambda, \alpha, \beta, X, y),$$

where the integration is over $\lambda, \alpha, \beta \geqslant 0$ and all $X \in R^{k\times k}$ and $y \in R^k$. Denote the ith row of X by $(X^i)^{\mathrm{T}}$. The exponential factor simplifies as follows.

$$\begin{aligned}
&\|M_2^*(\alpha, \lambda, \beta)^{\mathrm{T}}\|^2 + \|X\|^2 + \|y\|^2 \\
&\quad = \|-\lambda \mathbf{1}_k + X\alpha\|^2 + (-\lambda + y^{\mathrm{T}}\beta)^2 + \|-\lambda\mathbf{1}_k - X^{\mathrm{T}}\beta\|^2 + \|X\|^2 + \|y\|^2 \\
&\quad = \sum_{i=1}^{k}\left[\left(X^i - \frac{\lambda\alpha}{1+\|\alpha\|^2}\right)^{\mathrm{T}} (\alpha\alpha^{\mathrm{T}} + I)\left(X^i - \frac{\lambda\alpha}{1+\|\alpha\|^2}\right)\right] + \frac{(k+1)\lambda^2}{1+\|\alpha\|^2} \\
&\qquad + \|X^{\mathrm{T}}\beta + \lambda\mathbf{1}_k\|^2 + (-\lambda + y^{\mathrm{T}}\beta)^2.
\end{aligned}$$

Denote by $S(X)$ the absolute value of the determinant of X. We know that

$$\mathscr{E}[S(y, X)] = \sqrt{\frac{2^k}{\pi(1+\|\alpha\|^2)}}\,\Gamma((k+1)/2)$$

and

$$\sigma[S(y, X)] = \sqrt{\frac{k! - (2^k/\pi)(\Gamma((k+1)/2))^2}{1 + \|\alpha\|^2}}.$$

As in the previous sections,

$$\mu = \frac{\lambda\alpha}{\sqrt{1 + \|\alpha\|^2}}.$$

Next, let

$$\psi(y, X) = \psi_{\lambda,\alpha}(y, X) = \frac{|\det R|^{-k/2}}{(2\pi)^{k(k+1)/2}} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{k} (X^i - \mu)^T R^{-1}(X^i - \mu) + \|y\|^2\right\},$$

which is equal to the joint probability density function of the variates $X^1, \ldots, X^k$ and y assuming the X^i's are drawn independently from a k-normal distribution with mean vector μ and variance-covariance matrix $R = (\alpha\alpha^T + I_k)^{-1}$ and y is drawn (independently) from the standard k-normal distribution. We next define a matrix $Q(y, X)$ as in the case of $V_4(m, n, k)$. As a matter of fact, the details of the analysis from this point and on have already appeared in one of the previous cases. It follows that as in the case of $V_4(m, n, k)$, the contribution of this type of matrices to the expected number of steps, tends to zero when n tends to infinity while m is fixed, that is

Theorem 5.2. *For every fixed m and k,*

$$\lim_{n\to\infty} n^k V_2(m, n, k) = 0.$$

6. Conclusion

We have established in this paper that $\rho(m, n)$ is bounded from above by a function of m. Recall that $\rho(m, n)$ reflects the average number of steps and not the actual running time. The time it takes to perform a single step is in the worst-case proportional to n. Thus, the average running time does increase with n but is, asymptotically, linear in n. This may not be a surprise, in view of the existence of (worst-case) linear time algorithms for every fixed m, as shown by the author elsewhere [16].

It may be argued that the result of this paper is due to the fact that, under Smale's model, as n tends to infinity (with m fixed) the problem is unbounded with probability tending to one. The latter is of course true but does not explain the fact that the average number of steps of the self-dual algorithm tends to a finite limit.

To understand this claim, consider the following linear programming problem:

$$\begin{aligned} \text{Maximize} \quad & c^{\mathrm{T}}x \\ \text{subject to} \quad & Ax \leq b, \\ & x \geq 0 \end{aligned}$$

(where $x, c \in R^n$, $A \in R^{m \times n}$ and $b \in R^m$). Let us work (for example) with a weaker probabilistic model in which the coefficients are drawn independently from any distributions, provided each is positive or negative with equal probabilities. Consider a fixed column j ($1 \leq j \leq n$). If $s_j > 0$ and $a_{ij} < 0$ for $i = 1, \ldots, m$, then the problem is both feasible and unbounded, since for any set of values for the other variables, we can select a sufficiently large value of x_j that will satisfy all the constraints and will let us increase the objective-function value indefinitely, even when the rest of the variables are fixed. In such a case we say that j is a *good* column. Under our model, the probability that the jth column will be good is equal to $2^{-(m+1)}$. Thus, an efficient algorithm can be designed as we argue below.

The algorithm first scans the columns, one after the other, to check whether any of them has the sign pattern that creates unboundedness, in which case it stops (declaring the problem feasible and unbounded, and presenting the discovered good column as evidence). If none of the columns is good, then the algorithm proceeds like a variant of the simplex algorithm. Now, the probability that a good column exists is equal to

$$1 - \left(1 - \frac{1}{2^{m+1}}\right)^n,$$

which tends to one whenever n tends to infinity while m is fixed. Moreover, the *expected* number of columns we need to check before we discover a good one (or recognize that none is good) is less than 2^{m+1}, independently of n. Thus, for n sufficiently large, the algorithm will, very probably, discover a good column within a number of steps of order 2^{m+1}. However, this does not say much about the *expected* number of steps. Notice that the expected number of steps depends on what happens in the rare event in which none of the columns is good. However, any non-cycling simplex algorithm cannot perform more than $\binom{m+n}{m}$ steps. It follows that the contribution of this number to the average is smaller than

$$\left(1 - \frac{1}{2^{m+1}}\right)^n \binom{m+n}{m}$$

which tends to zero when n tends to infinity while m is fixed. Thus, such an analysis predicts an excellent asymptotic expected performance under the present model. On the other hand it pertains to a different algorithm and does not explain the particular behavior of the self-dual method.

At this point we can argue that the computational experience in linear programming to date may not have shown us the theoretical expected number of steps, but rather the performance in the vast majority of the cases. It may well be that the expected number of steps is exponential as a function of two variables (see [17] for an interesting related analysis). The observed phenomenon that the number of steps is usually less than $3m$ ([12]) does not necessitate that the limit of $\rho(m, n)$, when n tends to infinity, is a linear function of m. It may well be exponential in m.

Another aspect of this argument is that it may be the *conditional* expected number of steps, given that the problem is bounded, grows to infinity with n, even when m is fixed. We note that the growth, in terms of n when m is fixed, is polynomial since it is bounded by $\binom{m+n}{m}$. On the other hand, it follows from our analysis in this paper that the conditional number, $\rho^c(m, n)$, satisfies

$$\rho^c(m, n) \leq C(m)\left(\frac{1}{1-(1/2^{m+1})}\right)^n,$$

where $C(m)$ is such that $\rho(m, n) \leq C(m)$. So all we can say at this point is that $\rho^c(m, n)$ is also bounded by an exponential in terms of n (when m ix fixed), where the base of the exponent approaches 1 rapidly with m.

Finally, since the first version of this paper was written, lexicographic variants of the self-dual method have been noticed to perform on the average no more than $O(m^2)$ pivot steps for any n [3, 4, 5, 6, 22]. Adler and Megiddo [5] even proved a quadratic lower bound under certain conditions. However, it is still not known whether the starting point e yields a polynomial expected number of steps.

Acknowledgements

The author is indebted to Steve Smale for the discussions and the exchange of notes he had with him on the subject. The paper by Craig Tovey and Gideon Weiss [23] was of considerable help in correcting an error in an earlier version of this paper. The author is also thankful to Ilan Adler and Ron Shamir for their observations on the model as reflected in part of the conclusion of the paper. Finally, special thanks are due to an 'anonymous' referee whose comments were of great value during the preparation of the revised form of the paper.

References

[1] I. Adler, "The expected number of pivots needed to solve parametric linear programs and the efficiency of the Self-Dual Simplex method", Technical Report, Department of Industrial Engineering and Operations Research, University of California, (Berkeley, CA, June 1983).

[2] I. Adler, R.M. Karp and R. Shamir, "A family of simplex variants solving an $m \times d$ linear program in expected number of pivot steps depending on d only", Report UCB CSD 83/157, Computer Science Division, University of California, (Berkeley, CA, December 1983).

[3] I. Adler, R.M. Karp and R. Shamir, "A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of steps", Report UCB CSD 83/158, Computer Science Division, University of California (Berkeley, CA, December 1983).

[4] I. Adler and N. Megiddo, "A simplex-type algorithm solves linear programs of order $m \times n$ in only $O((\min(m, n))^2)$ steps on the average", preliminary report, November 1983.
[5] I. Adler and N. Megiddo, "A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension", in: *Proceedings of the 16th Annual ACM Symposium on Theory of Computing* (ACM, New York, NY, May 1984) pp. 312-323. Also, *Journal of the Association for Computing Machinery* 32 (1985) (to appear).
[6] I. Adler, N. Megiddo and M.J. Todd, "New results on the average behavior of simplex algorithms", *Bulletin of the American Mathematical Society* 11 (1984) 378-382.
[7] C. Blair, "Random linear programs with many variables and few constraints", Faculty Working Paper No. 946, College of Commerce and Business Administration, University of Illinois at Urbana-Champaign, (Urbana, IL, April 1983).
[8] R.G. Bland, "New finite pivoting rules", *Mathematics of Operations Research* 3 (1978) 103-107.
[9] K.-H. Borgwardt, "Untersuchungen zur asymptotik der mittleren schriftzahl von simplexverfahren in der linearen optimierung", Dissertation, Universität Kaiserlautern (1977).
[10] K.-H. Borgwardt, "Some distribution-independent results about the asymptotic order of the average number of pivot steps of the simplex method", *Mathematics of Operations Reserach* 7 (1982) 441-462.
[11] K.-H. Borgwardt, "The average number of steps required by the simplex method is polynomial", *Zeitschrift für Operations Research* 26 (1982) 157-177.
[12] G.B. Dantzig, *Linear programming and extensions* (Princeton University Press, Princeton, New Jersey, 1963).
[13] M. Haimovich, "The simplex algorithm is very good! - On the expected number of pivot steps and related properties of random linear programs", Technical Report, Columbia University, (New York, NY, April 1983).
[14] G. Kolata, "Mathematician solves simplex problem", *Science* 217 (1982) 39.
[15] C.E. Lemke, "Bimatrix equilibrium points and mathematical programming", *Management Science* 11 (1965) 681-689.
[16] N. Megiddo, "Linear programming in linear time when the dimension is fixed", *Journal of the Association for Computing Machinery* 31 (1984) 114-127.
[17] N. Megiddo, "On the expected number of linar complementarity cones intersected by random and semi-random rays", *Mathematical Programming* 35 (1986) 225-235.
[18] N. Megiddo, "A note on the generality of the self-dual simplex algorithm with various starting points", in: *Methods of operations research* (Oelgeschlager, Gunn & Hain, 1985), to appear.
[19] L. Santalo, *Integral geometry and geometric probability* (Addison-Wesley, Reading, MA, 1976).
[20] S. Smale, "On the average number of steps of the simplex method of linear programming", *Mathematical Programming* 27 (1983) 241-262.
[21] S. Smale, "The problem of the average speed of the simplex method", in: A. Bachem, M. Grötschel and B. Korte, eds., *Mathematical programming: The state of the art* (Springer-Verlag, Berlin, 1983) pp. 530-539.
[22] M.J. Todd, "Polynomial expected behaviour of a pivoting algorithm for linear complementarity and linear programming problems", Technical Report No. 595, School of Operations Research and Industrial Engineering, Cornell University, (Ithaca, NY, November 1983), *Mathematical Programming* 35 (1986) 173-192.
[23] C.A. Tovey and G. Weiss, "A note on the volumes of random simplices", ISyE Technical Report # J-84-8, Georgia Institute of Technology, (Atlanta, GA, June 1984).