Fast Calculation of Singular Values for MIMO Wireless Systems

Glenn Dickins and Leif W. Hanlen

***Abstract*— The standard (point-wise) linear channel model for MIMO wireless systems provides a simplistic mapping from antenna elements to continuous (operator) view point of wireless channels. Low-rank, high-dimension sampling matrices generated by Ray-Tracing may be used to estimate (with error) the "true" operator channel. In order to achieve reasonable estimation error bounds, intractably large dimension matrices must be used for Ray Tracing.**

We consider an algorithm for estimating the singular values of the large dimensional matrix via application of Power Factorisation. We show there is no preferential basis choice for preconditioning and provide a simple algorithm for high-speed evaluation of the dominant singular values.

***Index Terms*—MIMO Systems, Operators, Antenna arrays**

I. Introduction

The advent of multiple-input multiple-output (MIMO) wireless communications has promoted the concept of high bandwidth wireless systems employing large numbers of antenna elements at the transmit and receive ends of the wireless link. Much interest [1] has been devoted to closely spaced array elements and the effects of correlation for small wireless devices. Much of the wireless communication literature models antenna elements as points in space. We refer to such models as *point-wise models.*

Recent work [2–5] has shown that spatially diverse wireless systems may be more appropriately modelled using *continuous* spatial techniques, which focus on the continuous nature of space, rather than the individual antenna elements. In particular, these techniques consider the channel as an *operator*, [6], rather than a collection of discrete points. In [7] it was shown that antenna elements may be seen as "samples" of the continuous channel.

We shall focus on simple models of the form given in [8], where the antenna elements are assumed to be "dense." It has been shown [4] that the point-wise approximation of the continuous channel may be interpreted as a particular choice of orthonormal expansion, which miss-estimates the continuous mode connection strengths and requires a large number of sample points to provide reasonable accuracy. The size of the sampling matrix generated for estimating the operator channel must be extremely large and this gives a channel matrix of large dimension and low rank, with large computational complexity to calculate the channel singular values.

G. Dickins is with the Research School of Information Science & Engineering, the Australian National University, Canberra, ACT 0200, Australia. Email: glenn.dickins@anu.edu.au

L. Hanlen is with National ICT Australia, Canberra, ACT 0200, Australia, and affiliated with the Australian National University, Canberra, ACT 0200, Australia. Email: leif.hanlen@nicta.com.au

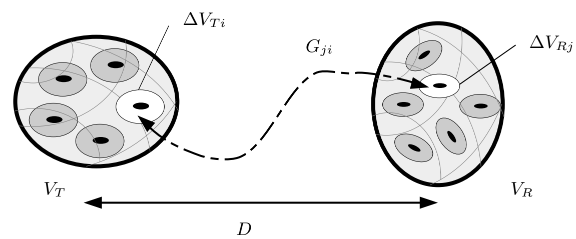


Fig. 1. Continuous solid regions V_T (transmit) and V_R (receive) containing t and r discrete antenna elements respectively. Two particular elements are highlighted, with the corresponding channel matrix entry G_{ji} shown.

This paper is arranged as follows: in section II we describe the channel model involved, and discuss decomposition of functions into particular choices of bases. Section III presents numerical comparisons of the ray tracing technique including computational complexity for both direct (point wise) estimation, and using the power factorisation method. We draw conclusions in section IV.

II. Channel Model

Consider the physical arrangement shown in Figure 1. We have shown two regions V_T and V_R in space. The region V_T generates signals which are transmit to V_R over a given channel. The region V_T contains t sources (transmit elements) which produce a field within V_R, and V_R contains r receive elements which sample the field within V_R. We have superimposed "discrete" antenna elements contained within each region, shown as black spots in Figure 1. Each (transmit/receive) element may be considered as the centroid of a given sub-region. The i^{th} transmit element is located at a point $\rho_{Ti} \in V_T$ and may be considered as being contained by the small sub-region ΔV_{Ti} as shown. Each sub-region ΔV_{Ti} is assumed to be disjoint and the collection of sub-regions occupy the whole of V_T, *i.e.*, $V_T = \sum_i \Delta V_{Ti}$. Similarly, the receive elements are contained within small disjoint sub-regions ΔV_{Rj} whose collective occupies all of V_R. The enclosing regions V_T and V_R are separated by a distance D (centre-to-centre). We shall assume that the channel is finite-frequency, has no inter-symbol interference and is time-invariant.

We shall collect the transmit and receive signals in vectors $x \in \mathbb{C}^{1\times t}$ and $y \in \mathbb{C}^{1\times r}$ respectively, to give the well-known MIMO linear channel model

$$y = Hx + w \tag{1}$$

where the transfer matrix $H \in \mathbb{C}^{t\times r}$ gives the complex channel gain coefficients between points in V_T and points in V_R.

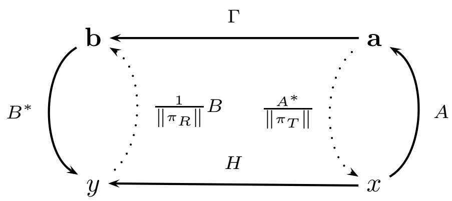


Fig. 2. Duality between discrete and continuous domains in noise free environment. Reversibility of arcs is dependent upon "sufficiently" dense sampling criterion [7] and equal-norm sampling functions.

If we consider the (transmit) signal $\psi(\mathbf{r}_T)$ generated by x in V_T, and the (receive) signal $\phi(\mathbf{r}_R)$ which is produced in V_R and sampled to give y we may write $\psi(\mathbf{r}_T)$ and $\phi(\mathbf{r}_R)$ in terms of complete, orthonormal sequences $\{\varphi_i(\mathbf{r}_T)\}_{i=1}^{\infty}$ and $\{\vartheta_j(\mathbf{r}_R)\}_{j=1}^{\infty}$. We have

$$\psi(\mathbf{r}_T) = \sum_i a_i \varphi_i(\mathbf{r}_T) \quad a_i \triangleq \langle \psi(\mathbf{r}_T), \varphi_i(\mathbf{r}_T) \rangle \tag{2}$$

$$\phi(\mathbf{r}_R) = \sum_j b_j \vartheta_j(\mathbf{r}_R) \quad b_j \triangleq \langle \phi(\mathbf{r}_R), \vartheta_j(\mathbf{r}_R) \rangle \tag{3}$$

and we may collect the elements a_i and b_j in vectors: $\mathbf{a} = \{a_1, \ldots\}$, $\mathbf{b} = \{b_1, \ldots\}$ and write:

$$\mathbf{b} = \Gamma \mathbf{a} \tag{4}$$

If we assume that each transmit antenna element j has an orthonormal spatial signature $\mathbf{s}_{Tj}$ then we may write the transmit vector in terms of a sampling matrix [7, 9] A,

$$\psi(\mathbf{r}_T) = Ax \quad A_{ji} = \langle \mathbf{s}_{Tj}, \varphi_i(\mathbf{r}_T) \rangle \tag{5}$$

and similarly, for the receive signal $\phi(\mathbf{r}_R)$, we may assume each receive antenna element j has an orthonormal spatial signal $\mathbf{s}_{Rj}$ and use a sampling matrix B to give y

$$y = B^* \phi(\mathbf{r}_T) \quad B_{ji} = \langle \mathbf{s}_{Rj}, \vartheta_i(\mathbf{r}_R) \rangle \tag{6}$$

Combining (1), (4), (5) and (6) we have:

$$y = B^* \Gamma A x + \hat{w} \tag{7}$$

Figure 2 describes the duality of the continuous and discrete models given by (7). The emphasis of previous work [7] was in generating H given Γ. In this paper we shall not be interested in noise or the capacity of the channel: our motivation is to examine the properties of H and Γ.

It has been shown [9] that the transfer matrix Γ is inherently finite rank, and this rank is given by the parameters of the multipath channel. Further, it has been shown [4] that in order to accurately estimate the channel Γ from H using point-like sampling *the number of points required is large.* Point-like sampling (without any prior channel preconditioning) suffers from two serious flaws:

1) The number of points is extremely large (giving a high-dimension matrix H) while the rank of the channel is finite (the rank of H is small)
2) The error associated with the estimate $\widehat{\Gamma}$ given H is unbounded. In particular, the trace of the estimate matrix $\widehat{\Gamma}$ remains fixed, which prevents a natural measure of convergence.

We shall address both issues below. Our first step is to develop a means to reduce the computational complexity of the singular value estimation, given that the dimension of the matrix H is much larger than the number of non-negligible singular values. We next examine the convergence characteristics of the algorithm, and show that the convergence is monotonically increasing.

III. Reducing Computational Complexity

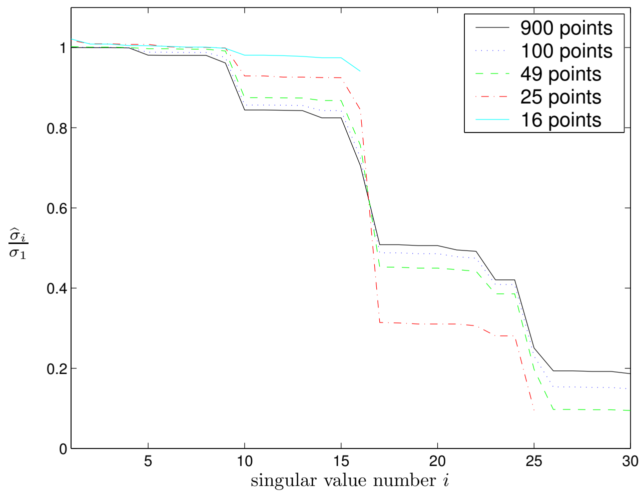
A singular value decomposition (SVD) is required to estimate the channel singular values and decompose the continuous channel into independent parallel channels. Where the transmission function has been appropriately sampled, there will be a large number of points over the transmit and receive volumes, thus requiring a large matrix SVD. Given an equal number of transmitter and receiver sample points, n, computation of the SVD requires $O(n^3)$ [10]. From an estimation view-point, it is desirable to ensure that n is sufficiently large to capture all degrees of freedom from the channel. In most cases the number of non-negligible singular values will be much less than n, thus we know that the transmission matrix may be approximated by a low rank matrix.

A. Truncated Basis Approximation

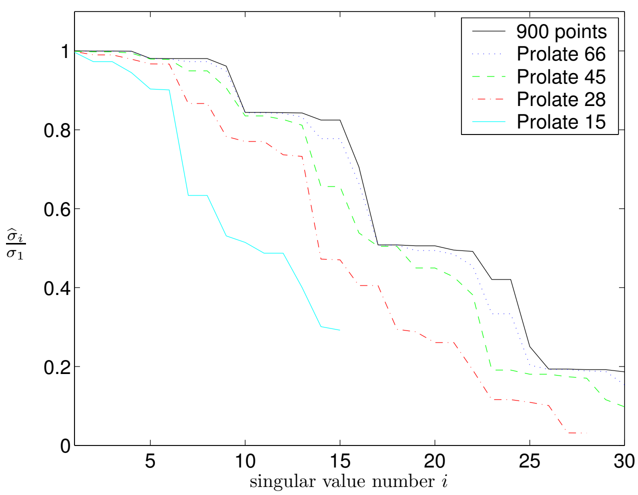
An natural method to generate a low dimension matrix $\widehat{\Gamma}_k$ whose singular values are equal to the largest k singular values of Γ is to apply the transformation of Figure 2: If we choose the basis for the channel, and apply the basis functions to our sampling points in the form of matrices A and B, a lower dimension approximation of the matrix Γ may be formed. Such a basis would well approximate the most likely channel eigenfunctions using only a truncated subset of the basis. To approximate the channel singular values, it is then sufficient to calculate the singular values of the truncated basis projection of the transmission matrix.

Given unitary $n \times r$ matrices for the transmitter volume B_T and receiver volume B_R the $r \times r$ channel matrix approximation would be $B_R{}^* G B_T$. To calculate this matrix and determine its singular values would require $4n^2 r + 4r^3$ operations. This approach would become more computationally efficient when $n/r \gtrsim 1.46$ and $O(n^2)$ for $n \gg r$.

This approach relies on the existence and invariance of such a basis function set. For simple geometries, where the regions are flat and parallel, it can be shown that a low order Prolate type basis function set across the region gives a good approximation for the channel singular values. In this case, a monotonically increasing approximation of the singular values is obtained providing a more stable estimate than a sparsely sampled transmission matrix. Similar properties hold for other simple geometric arrangements: such as using Spherical Harmonics for communication between concentric spherical shells.

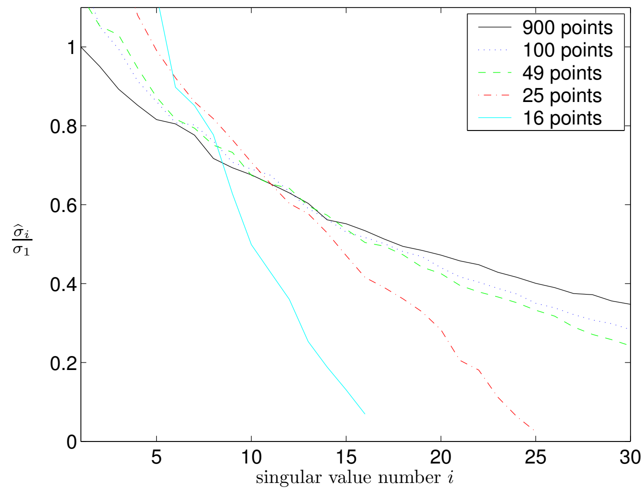


(a) Reduced point set (note over-estimation)

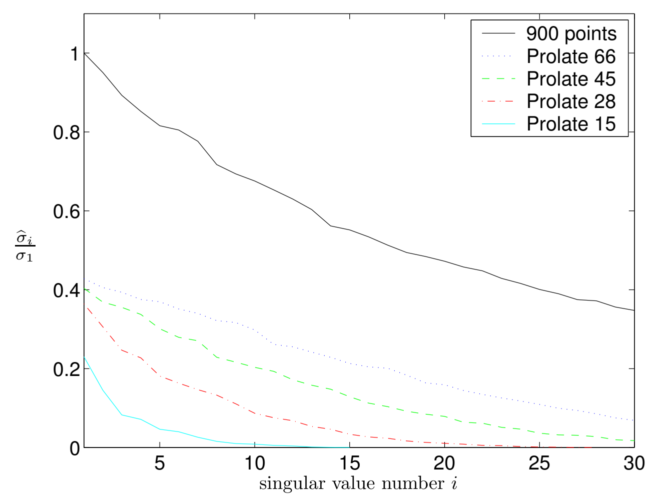


(b) Truncated basis estimation (note monotone convergence)

Fig. 3. Approximation of singular values via SVD on reduced point set and SVD on basis projection.



(a) Reduced point set (note over-estimation)



(b) Truncated basis estimation (note poor convergence)

Fig. 4. Approximation of singular values via SVD on reduced point set and SVD on basis projection for 10 reflective scattering bodies.

However, for unknown channels without a natural geometry[1] there is no guarantee that a particular basis may used independently of the channel. Figure 3 compares the estimates of singular values with truncated basis approximation. Figure 3(a) shows the overestimated singular values when reduced point sets are used. Figure 3(b) demonstrates the monotonic convergence to the actual singular values when a truncated basis is applied to the large sampled channel matrix. This figure is for a configuration of two flat planes of 20 wavelengths separated by a distance of 100 wavelengths.

When the system is perturbed by angular rotation, or by introducing several scattering objects, it can be seen that the basis functions are no longer appropriate for estimating the singular values. Figure 4 introduces to the configuration 10 scatterers having a 45° standard deviation in angle. Figure 4(a) provides the truncated point method result while Figure 4(b) shows that the prolate basis now provides a poor approximation. Similar result can be shown for other basis functions such as a Fourier basis, scaled spherical harmonics, and Gabor wavelets.

Although this shows that some general basis types are not appropriate, it does not eliminate the existence of a suitable basis. To investigate this we calculated the dominant eigen-modes for a sample of channel configurations. Principle component analysis, also known as Karhunen-Loève expansion, was then used to search for the presence of common eigen modes which would provide a suitable basis for general truncation. The 10 dominant modes from each of 100 configurations of the channel were obtained. The channels were statistically created with a Gaussian angular diversity about a perpendicular direction of arrival. Figure 5 shows the singular values of the matrix consisting of these 1000 channel transmission modes. It can be seen that as the angular diversity increases the singular value spectrum becomes flat. Thus there is no common eigen-modes and consequently no preferred basis for the possible channel eigenmodes as the arrival angle becomes uncertain. A reduced order basis is only feasible with a preferred direction of arrival and low angular diversity. Although the dominant amplitude pat-

[1] If we are reduced to estimating the channel, it is unrealistic to imagine it will have a simple geometric arrangement!

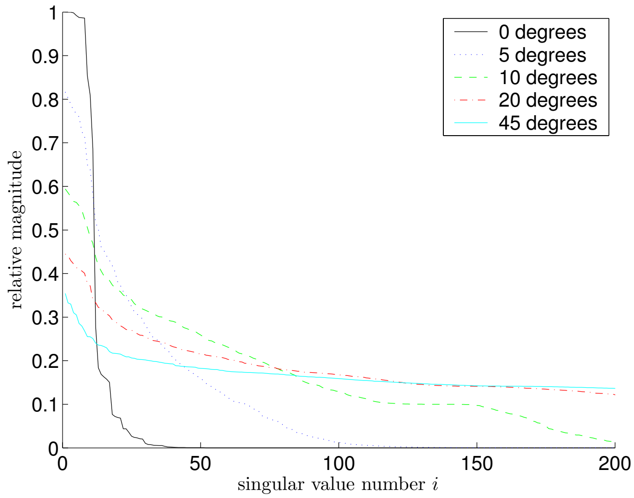


Fig. 5. Demonstration of lack of invariant basis function. As the scattering diversity increases, the span of eigenfunctions increases - there can be no preferred basis with high angular diversity

terns across the region remain largely unchanged, the differing directions of arrival introduce a variable phase offset across the receiver volume. In a rich scattering environment, the dimensionality of this potential variation approaches critical sampling of the region and no basis would offer any computational advantage. Unless a single dominant direction of arrival is identified and compensated for, truncated basis approximation does not offer any benefits over simply sampling at high density.

B. Power Factorisation

In general we know that the spectrum of the channel matrix will be concentrated, with few large singular values compared to the number of samples required for critical sampling of the receive and transmit volumes. We desire a method to extract the significant singular values from a critically sampled channel matrix without incurring the expense of a complete singular value decomposition. To achieve this we introduce the process of power factorisation [11, 12] based on the method of orthogonal iteration for determining dominant eigenvalues [10].

Given the $n \times n$ channel matrix, G, and an estimate for the critical number of singular values, r, we first construct an initial random matrix A_0 of size $n \times r$ and then iterate

$$\begin{aligned} B_k &= G^* A_{k-1} N_k \\ A_k &= G B_k \end{aligned} \tag{8}$$

where N_k is a matrix selected at each step to ensure that B remains unitary (Gram-Schmidt or QR decomposition). This provides a rank r factorisation that converges to the global minimum of $\|M - AB^*\|_F$ [11] under the Frobenius norm. After several iterations, the significant singular values of G can be estimated from a SVD of A_k. The computational requirements for the k iterations and the SVD are given in Table I.

It can be shown that for $n/r \gg 1$, the Power Factorisation method is more efficient than direct SVD if $k < n/(2r)$. Figure 6 shows that convergence of the singular values for $r = 20$ is achieved in three to four iterations. In this case $n = 900$, thus a significant computational saving is possible.

TABLE I

Computational complexity of Power Factorisation method

Iteration	Complexity
$G^* A_{k-1}$	$2n^2 r$
Gram-Schmidt N_k	$2nr^2$
GB_k	$2n^2 r$
SVD(A)	$2nr^2 + 2r^3$
TOTAL	$4kn^2 r + 2(k+1)nr^2 + 2r^3$

C. Convergence

Given the optimal rank r approximation of M as $\widehat{M}$ then the approximation AB^* will converge as

$$\left\| \widehat{M} - AB^* \right\|_F \leq C \left(\frac{\sigma_{r+1}}{\sigma_r} \right)^{2k} \tag{9}$$

for some constant C. Due to the compact nature of the channel operator Γ, the singular values of H are bounded [6]. Further the singular values may be ordered such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k \geq \cdots \geq 0$. For any value $\epsilon > 0$, there is a number N_c such that for $i \geq N_c$, $\sigma_i < \epsilon$ [6]. Convergence would be slow for an *i.i.d.* channel matrix, however this is unlikely given sufficiently dense sampling. Figure 7 shows the convergence of the norm compared with the bound for $r = 20$. The most significant reduction in the norm is obtained in the first few iterations.

Figure 7 also shows the average relative error in the singular values calculated as,

$$\frac{1}{r} \sum_{i=1}^{r} \frac{(\widehat{\sigma}_{ik} - s_i)^2}{\sigma_i} \tag{10}$$

where $\widehat{\sigma}_{ik}$ is the estimate of the i^{th} singular value from the k^{th} iteration. The average relative error in the singular values is less than 2% in both cases after only four iterations. Since the larger singular values converge faster (*e.g.*, Figure 6), the absolute error is quite small after only a few iterations.

Consider the expansion of the second step in the iteration, with

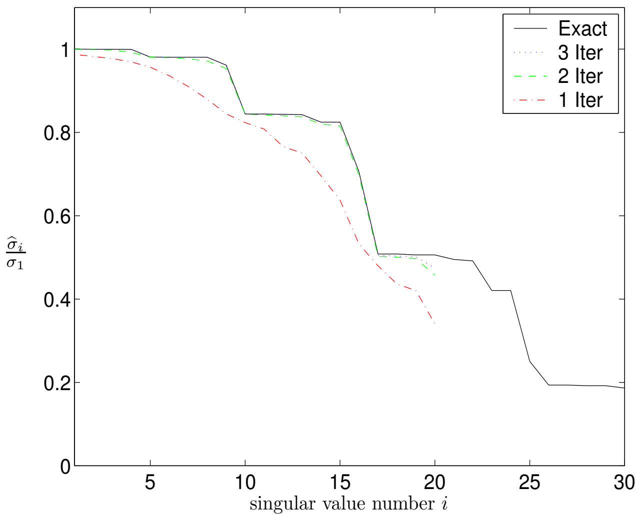
$$G = UDV^* \tag{11}$$

i.e., U, D and V represent the true singular value decomposition of G. Then we may write

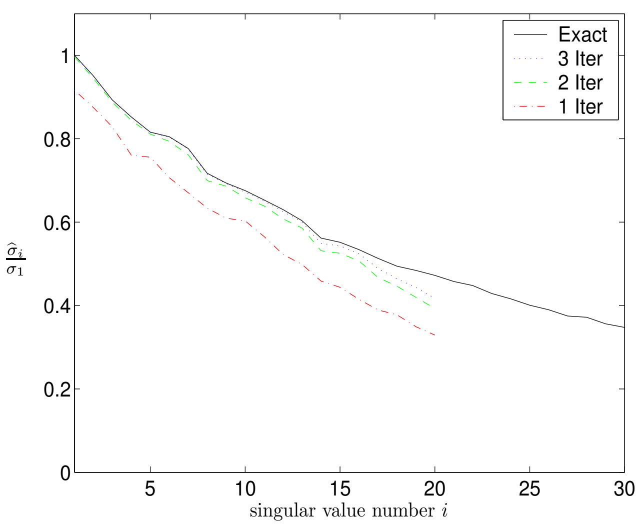
$$A_k = GB_k = UDVB_k \tag{12}$$

Since the r largest singular values of G, reside in the upper r quadrant of D and (VB_k) is necessarily unitary, the singular value estimates increase monotonically as VB_k ultimately converges to an upper identity [2]. This avoids the problem of overestimating singular values as seen when the receive and transmit volumes are under sampled. The span of B converges to cover the truncated span of the true channel excitation modes. Although complete convergence can be slower, the energy in the actual dominant eigen-modes orthogonal to the span of B quickly converges to zero.

[2]Convergence to the upper identity requires a unique solution, *i.e.*, distinct singular values. If some values are repeated, the iterations will converge so that singular vectors remain within the same singular space.

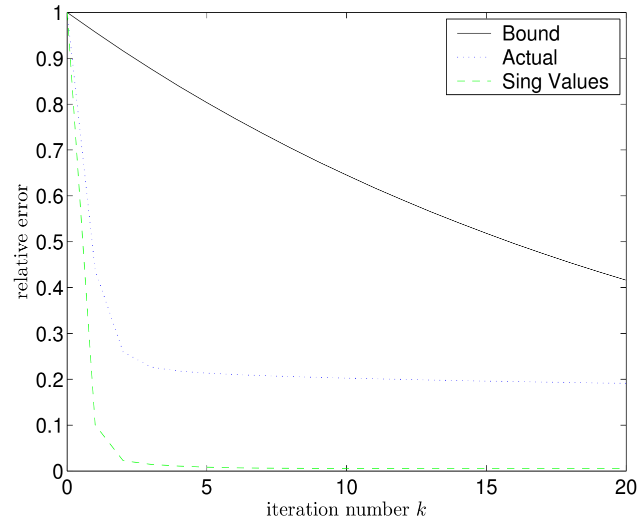


(a) Simple (direct) geometry

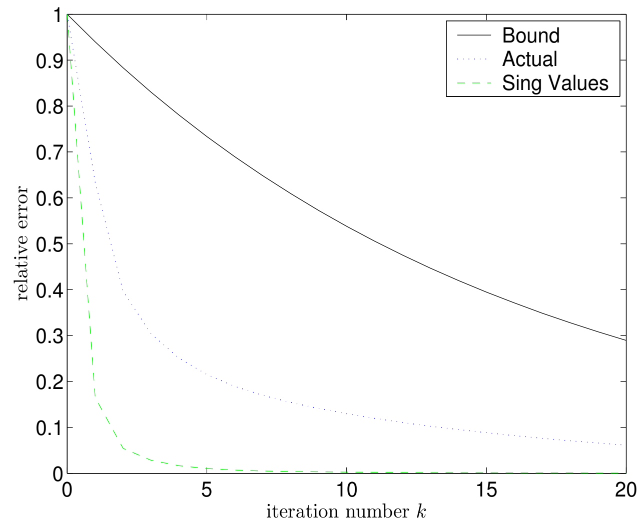


(b) Scattering environment

Fig. 6. Demonstration of power factorisation. The approximated singular values ($\widehat{\sigma}_{ik}$) rapidly and monotonically converge to the actual singular value (σ_i). In this example $r = 20$.



(a) Simple (direct) geometry



(b) Scattering environment

Fig. 7. Convergence of power factorisation. The actual Frobenius norm error is well within the bound for the early iterations. The RMS error in singular values rapidly decreases as the larger singular values are quickly approximated.

D. Implementation and Performance

Table II shows the cpu time requirements for calculating and estimating the singular values for various size channel matrices. The estimates are computed using the direct singular value decomposition, along with 2, 5 and 10 iterations of the power factorisation method for a rank 20 approximation. The simulation was implemented in MATLAB^{tm}and run on a Pentium–IV ^{tm}. For large numbers of points, the proposed approach offers an order of magnitude improvement in time even allowing for 10 iterations of the power factorisation.

Note that further investigation is proposed to compare this computational efficiency to other methods of selective singular value calculation as outlined in [13–15] and implemented in the SVDPack library [16]. It may also be possible to improve the convergence of the power series method by using a technique similar to Ritz acceleration for orthogonal iteration [10].

IV. CONCLUSION

We have shown that there is no preferential pre-conditioning sampling basis which preserves the singular values and guarantees a low-dimension representation of the channel in general. If we restrict ourselves to particular geometries, then pre-conditioning may be of assistance.

We have described a simple Ray Tracing algorithm, Power Factorisation, for estimating the singular values of a continuous spatial channel. This technique allows us to exploit the simplicity of the point-wise modelling approach while overcoming the large number of samples required for point-wise modelling.

The power factorisation method has been shown to converge monotonically to the correct singular values, and is significantly less computationally expensive than direct SVD on the large point-wise channel matrix.

TABLE II

CPU time using Pentium–IV^{tm} processor, and MATLAB^{tm}. Fixed number of points comparing SVD and 2, 5 and 10 iterations of the Power Factorisation.

Points	SVD	2 Itns	5 Itns	10 Itns
1225	175.2	5.7	7.8	11.6
900	64.4	3.9	4.2	6.1
625	19.3	1.5	2.1	3.1
400	4.0	0.6	0.9	1.3
225	0.8	0.2	0.3	0.5

REFERENCES

[1] K. Yu and B. Ottersten, "Models for MIMO propagation channels: a review," *Wireless Communications and Mobile Computing*, vol. 2, pp. 653–666, Nov. 2002.

[2] T. Svantesson and J. Wallace, "On signal strength and multipath richness in multi-input multi-output systems," in *ICC03*, Anchorage, Alaska, May. 11–15 2003, pp. 2683–2687.

[3] R. Kennedy and T. Abhayapala, "Spatial concentration of wave-fields: Towards spatial information content in arbitrary multipath scattering," in *Proc. 4th Australian Communications Theory Workshop*, Melbourne, Australia, Feb. 2003, pp. 38–45.

[4] L. W. Hanlen, "Channel modelling and capacity analysis for multi-input, multi-output, wireless communication systems," Ph.D. dissertation, The University of Newcastle, June 2003.

[5] T. Pollock, "On limits of multi-antenna wireless communications in spatially selective channels," Ph.D. dissertation, The Australian National University, July 2003.

[6] E. Kreysig, *Introductory functional analysis with applications*. New York, USA: John Wiley & Sons, 1978.

[7] M. Fu and L. W. Hanlen, "Capacity of MIMO channels: A volumetric approach," in *ICC03*, Anchorage, Alaska, May. 11–15 2003, pp. 2673–2677.

[8] N. Chiurtu, B. Rimoldi, and E. Telatar, "Dense multiple antenna systems," in *IEEE Information Theory Workshop*, Cairns, Australia, Sept. 2–7 2001, pp. 108–109.

[9] T. Pollock, T. Abhayapala, and R. Kennedy, "Fundamental limits of MIMO capacity for spatially constrained arrays," in *Proc. 4th Australian Communications Theory Workshop*, Melbourne, Australia, Feb. 2003, pp. 7–12.

[10] G. H. Golub and C. F. van Loan, *Matrix Computations*, 3rd ed. Baltimore, Maryland, USA: Johns Hopkins, 1996.

[11] R. Hartley and F. Schaffalitzky, "PowerFactorization: 3d reconstruction with missing or uncertain data," *Submitted*, 2003.

[12] T. Morita and T. Kanade, "A sequential factorization method for recovering shape and motion from image streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 8, pp. 858–867, 1997.

[13] G. H. Golub, F. T. Luk, and M. L. Overton, "A block lanczos method for computing the singular values and corresponding singular vectors of a matrix," *ACM Transactions on Mathematical Software*, vol. 7, no. 2, pp. 149–169, June 1981.

[14] S. V. Huffel, "Partial singular value decomposition algorithm," *Journal of Computational and Applied Mathematics*, vol. 33, no. 1, pp. 105–112, December 1990.

[15] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. V. der Vorst, Eds., *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*. Philadelphia: SIAM, 2000, http://www.cs.utk.edu/ dongarra/etemplates/.

[16] M. Berry, "Large scale singular value calculations," *International Journal of Supercomputer Applications*, no. 6, pp. 13–49, 1992.