Developments in Sensor Array Signal Processing

Prof J G McWhirter FRS FREng

QinetiQ Ltd, Malvern

Slide 1 (Title Slide)

I would like to start by thanking Graham and the organising committee for inviting me to give the plenary talk at this year's Asilomar conference. I feel very honoured to have been given this privilege and opportunity.

Slide 2 (Overview of Talk)

First of all let me give you a quick outline of my presentation. My choice of topics has had to be very selective, so this must be regarded as an overview and not a tutorial. I intend to focus on the topic of sensor array signal processing since that has been my main area of research for the last 20 years or so. I also believe that sensor array signal processing is now situated at the very forefront of research into digital signal processing more generally.

I want to start with a brief historical overview. Then I will go on to discuss some recent developments and current trends. In particular, I want to draw attention to an important trend from adaptive beamforming to blind signal separation and from second order statistics to higher order statistics. I would also like to point out the convergence which I see between this area and that of artificial neural networks. I will then go on to outline some of my current research and conclude by pointing out some future challenges.

Slide 3 (Sensor Array Signal Processing)

Sensor array signal processing, like DSP in general, has recently come of age - enabled by the revolution in digital processing technology. It has also been fuelled by some impressive research results. Only ten years ago we were struggling to persuade our military customers that this was an important topic capable of delivering practical benefits and that the necessary digital processing capability would soon be available. Now, thanks to mobile telephones, and other IT products which are so heavily based on embedded signal processing, our customers believe that the revolutionary capabilities offered by sensor array signal processing are realistic and affordable.

Sensor array signal processing has a very broad range of important applications. For example, it is widely believed to be the key to improving mobile telephone systems. The adaptive beamforming concepts I will be talking about later are now being exploited in this context to provide spatial diversity. They are also fundamental to the most recent research into multiple input, multiple output (or MIMO) communication channels. I believe that sensor array signal processing could revolutionise the design of future radar systems. Indeed research has been carried out by our group in Malvern into the concept of ‘element digitised array radar’ where the degree of freedom offered by advanced algorithms virtually allows us to take a clean sheet of paper in thinking about the design of future radar systems. Also, by way of contrast, sensor array signal processing is now beginning to play an important role in multiple sensor ECG and EEG measurements for medical diagnostics. I will be discussing the application to ECG later.

Slide 4 (Adaptive Null Steering)

In order to illustrate the sort of thing we can do with sensor array signal processing, I will say a little bit about the concept of adaptive null steering. This serves to illustrate why sensor arrays are so important. If you consider something like a phased array radar where the outputs of multiple identical sensors are combined electronically, you have many more degrees of freedom and can do things that weren't possible previously. In particularly you can adaptively design the angular response of a beam to steer a very low gain null in the direction of unwanted interference, whilst preserving the ability to look at any objects of interest. This would, of course, be physically impossible with the type of parabolic reflector used in a conventional radar antenna.

Slide 5 (Adaptive Beamforming)

So how can this capability be achieved in practice? This slide shows a very simplified schematic of such a system. The Y-shaped symbols represent the multiple sensors. Radiation arrives from various sources and it is assumed that the signal in each channel can be multiplied by a complex weighting factor. The complex weight serves to modify the phase and amplitude of the signal. When we combine the signals, we add them together with phase and amplitude values designed to achieve the output we want – that is, the desired signal with greatly reduced interference. It turns out that in most circumstances of interest, getting rid of the interference - assuming that it is uncorrelated with the desired signal - is equivalent to minimising the total output power from the array.

Now the output signal may be expressed mathematically as the scalar product of the complex weight vector and the vector (or snapshot) of samples across the array. In mathematical terms, we strive to minimise the power of this signal, subject to a linear constraint of the form shown on the slide. The constraint ensures that if a signal arrives with the type of wavefront expected from direction *theta*, (the chosen look direction), the gain of the array for that signal is maintained at a constant value.

Slide 6 (Least Squares Solution)

This leads directly to a form of least squares problem. In practice, we minimise an estimate of the output power, defined as the sum over all samples up to the present time of the modulus squared of the output of the array. Using some very simple maths, this leads to an expression of the form shown on the first line of this slide where $\mathbf{M}(n)$ denotes the familiar covariance matrix. Minimising this quantity subject to the linear constraint, constitutes a classical constrained least squares problem whose solution is given by the Gauss normal equations as shown here. The elements of the covariance matrix are just corresponding sample estimates of the covariance between signals from different channels so it can be seen that the solution only involves conventional second order statistics. An important point to note here, is that computing the weight vector $\mathbf{w}$ requires the solution of a set of p linear equations where p denotes the number of sensor elements. The number of arithmetic operations required for that procedure scales as p^3 - or as p^2 per sample time if we use a recursive updating algorithm. At the time when I first became interested in these techniques, this level of computation was well beyond the scope of real time digital processing – even using the best available technology. We were in a situation where we knew what we ought to do, but it was not technically feasible. That's what led to predominant use of the so-called least mean squares or LMS algorithm in the early 1980s. I would like to say a little bit about that algorithm now.

Slide 7 (LMS Algorithm)

In connection with the LMS algorithm, you may be more familiar with the type of problem shown here. The output to be minimised constitutes a linear combination of data samples represented by the vector **x***(t)* plus a so-called primary or reference signal *y* with unit gain. In this case, the weight vector **w** is not subject to any other constraint. This is the classical structure which arises, for example, in adaptive equalisation - the type of problem for which the LMS algorithm was originally developed. The LMS algorithm tries to minimise, in a statistical sense, the expectation of the output energy as shown on the slide. This is not accomplished by solving the least squares problem, but by taking various short cuts and approximations to derive a stochastic gradient update - the new weight vector is obtained from the old weight vector by simply adding a constant times the old data vector multiplied by the output produced at that time. It's so simple you might not believe it can work - but it does, and it's still one of the most widely applied algorithms for adaptive filtering and adaptive beamforming - a vital landmark for anyone working in this area.

The LMS algorithm involves minimal computation – in fact, the number of arithmetic operations per sample time scales linearly with *p*. Unfortunately It can be very slow to converge - the convergence rate is determined by the spread of eigenvalues of the covariance matrix **M** which is outside the control of the algorithm. Unless you can do something to reduce the eigenvalue spread, slow convergence will always be a potential problem with the LMS algorithm. Conversely, the least squares solution presented earlier does not suffer from the problem of slow convergence but requires much more computation.

Slide 8 (Canonical Problem and GSLC)

In my opinion, a particularly important development in the history of adaptive beamforming was the generalised sidelobe canceller (or GSLC) technique proposed by Griffiths and Jim in the late 1980s. In effect, the GSLC applies an exact mathematical transformation which converts the linearly constrained problem discussed earlier, into the canonical form shown here. This is, of course, the form which we assumed for the LMS algorithm so the GSLC serves to bridge the gap between these two methods.

In practice, you can choose any fixed beamformer and use it to form the output channel *y*. The data vector is also multiplied by the so-called blocking matrix **A** whose rows are chosen to span the null space of the vector **c**. In effect, each row of **A** guarantees a null in the desired signal direction. It turns out that any arbitrary weight vector **w** can be applied to the processed vector **x**, and the combined output, minimised w.r.t. **w**, is exactly the same as that obtained by solving the constrained problem.

Interestingly enough, when Terry Shepherd and I proposed the original adaptive beamformer based on QR decomposition, we did something equivalent. We derived a "constraint pre-processor" to transform the input data because our array was also designed to solve the canonical least squares problem. It turns out that our constraint pre-processor is a special case of the GSLC. I wish we had spotted the more general solution, since it is such a powerful and elegant mapping from the constrained to the canonical domain.

Slide 9 (QRD Processor Array)

Returning to the least squares computation, an important development occurred in the late 1970s or early 1980s. H T Kung published his seminal papers on systolic arrays. As far as real time matrix computations were concerned, systolic arrays provided some crucial light at the end of the tunnel at that time. The TRW single chip multiplier had just arrived on the scene and H T Kung showed how arrays of multiply-add components could be used to perform useful matrix computations. Essentially, Kung had shown how to avoid the Von Neumann bottleneck – a fundamental problem with the classic serial, or shared memory computer. At any rate, the systolic array concept led to a great surge of research on least squares algorithms of the type I described earlier. There was a strong emphasis on recursive update techniques, such as the RLS algorithm, pipelined Kalman filters and so on.

A significant development which I was involved with at the time, is the well-known triangular systolic array for solving the least squares problem in a recursive manner. It is based on the method of QR decomposition by Givens rotations which I won't discuss here since it is quite involved and rather old hat. The algorithm can be represented very effectively in terms of the underlying signal flow-graph as illustrated here. The signal flow-graph doesn't include any of the detailed timing or implementation details that we used to associate with systolic array design. It is a more abstract concept which represents the algorithm in a very precise way. It defines the sequence of operations which must be carried out in order to update the system from one time instant to the next, based on the latest snapshot of data. I haven't included any of the mathematics on this slide because those of you who already know it will be bored to tears and there isn't enough time to go into detail for those who are not familiar.

I was responsible for showing that the desired output from an adaptive beamformer can be obtained very simply from this array array without computing the optimum weight vector explicitly. The triangular array then operates like a super numerical filter, with signals and interference coming in at the top and a much cleaner signal emerging from the bottom. Not having to compute the weight vector obviously reduces the computational load and saves on hardware - but more importantly, it eliminates the need to solve a potentially ill-conditioned inverse problem and leads to a more stable algorithm.

Unfortunately, the weight vector itself is often required for other reasons, so the associated inverse problem can't always be avoided. For simplicity then, let me return to the basic least squares solution as defined by the Gauss normal equations. These can be solved quite comfortably nowadays using conventional serial computers which are fast enough for many applications.

Slide 10 (Unstabilised Beam Pattern)

The next slide shows the result obtained when this method was applied to computer simulated data from an experiment in which we modelled the effect of five separate sources of interference.

The yellow line plots the gain of the adapted beam as a function of angle. The red vertical bars indicate the direction of the interfering sources – the green bar indicates the specified look direction. You can see that the algorithm has done exactly what we asked it to do. It has directed deep nulls with very low gain towards the interference but held the gain at a much higher level in the look direction. Unfortunately, this type of angular response or beam pattern would be unacceptable to a radar systems designer. The gain of the array in other directions is much too high and high sidelobes like this cause serious problems in the presence of clutter. These high sidelobes are a direct result of the inverse problem being ill-conditioned. The weight vector has more degrees of freedom than the situation demands and the spare degrees of freedom can take on meaningless values determined only by noise and other random effects.

This aspect is illustrated quite dramatically by the following sequence, which shows how the beam pattern varies from one block of data to the next in the same experiment. This effect is often referred to as weight jitter. Random fluctuations like this are very bad news, even though the algorithm is doing its job – pointing the main beam where we want it and steering nulls towards the interference.

Slide 11 (Penalty Function Method)

David Hughes and I managed to solve this problem very effectively using a simple penalty function technique as outlined here. The radar designer is asked to specify the type of beam he or she would like to use under normal circumstances. This corresponds to a quiescent weight vector **w**q. An additional quadratic term, *E*bar, is then added to the least squares cost function. The purpose of this term, which is given a relatively low, positive weighting, is to penalise the adaptive algorithm for departing from the default beam pattern. This ensures that it will only do so when necessary to counteract strong interference, for which the cost is much higher. Since the penalty function term is quadratic in the weight vector, an analytic solution can still be obtained and the method is relatively easy to apply.

Slide 12 (Stabilised Beam Pattern)

The next slide shows the results we obtained, when the penalty function method was applied to the same set of data as before. The adapted beam has much lower sidelobes - consistent with the default beam - as well as deep nulls in the right direction. In the corresponding sequence, you can see that the effect of weight jitter has also been greatly reduced. The algorithm is now doing exactly what the radar designer wants – producing a stable beam with low side lobes, pointing it in the right direction and adaptively steering nulls towards the sources of interference wherever they happen to be.

Slide 13 (Sonobuoy Array)

I will conclude this part of my talk by showing some results we obtained using a stabilised adaptive beamformer to process real sonar data. This data was obtained from some underwater acoustic trials set up to test an experimental sonobuoy of the type shown here. This slide shows the sonobuoy being deployed at sea. It takes the form of an acoustic sensor array and you should be able to make out the individual sonar transducers on each of the five staves.

Slide 14 (Application to Sonar)

This slide shows the processed results, as they would be presented to a sonar operator. In each figure, the strength of the processed signal is plotted as a function of range along the vertical axis and bearing along the horizontal axis - the brighter the spot, the stronger the signal. The left-hand figure shows the result of using a conventional non-adaptive beamformer. The signal of interest is swamped by acoustic noise from a source which was located in the direction of the bright vertical bar and is seen in every range cell. The noise was received via the sidelobes as well as the main beam and this explains the other bright stripes. The right-hand figure shows the results obtained using a stabilised adaptive beamformer. The effect of the noise has been drastically reduced, and the object which we wanted to detect is now clearly visible as the brighter of the two remaining spots. When I first saw these results, which were produced by my colleague Cameron Speirs, I assumed that the other relatively weak spot was a false alarm. But on further investigation, we discovered that it was due to a shipwreck lying on the seabed – one that we knew nothing about a-priori. This really impressed the sonar experts, and has helped to ensure that adaptive beamforming is now taken very seriously in the underwater acoustics area. I could have shown results from radar or other areas of application but time does not permit me to do that in this talk.

Slide 15 (Blind Signal Separation)

I would now like to move on to the second phase of my talk – focussing on more recent developments. In particular I would like to talk about a technique known as blind signal separation (or BSS) which will serve to highlight some significant current trends. So what is blind signal separation? In the adaptive beamforming section, I assumed that the position of the sensor array elements was known, and that allowed me to point the main beam in a chosen direction. That in turn, allowed me to separate the signal of interest from various sources of interference. In blind signal separation, we assume that the position of the array elements is unknown, and that the array response cannot be calibrated in a meaningful way. As a result, it's impossible to form a beam which looks in any particular direction and the array is said to be blind – it is now just a collection of individual sensors.

Blind signal separation is a relatively new method, which avoids the need for a calibrated array and has been applied successfully to a wide range of problems including HF radio communications and foetal heartbeat analysis. It's based on an important new technique known as independent component analysis or ICA for short. ICA is a generic processing concept which uses higher order statistics and therefore requires the signals to be non-Gaussian. In other words, the probability distribution of the individual sample values must be non-Gaussian. This might appear to be a rather restrictive condition, but it is, in fact, a common property of sinusoids and other man-made signals such as those used for digital communications. In blind signal separation, the higher order statistics – in other words the cumulants of order greater than two - provide additional information, which compensates for the lack of calibration. In the case of Gaussian signals however, the higher order cumulants are known to be zero, so there is no extra information and the method cannot be applied.

Slide 16 (Blind Signal Separation)

This slide contains a simple schematic diagram of the set-up associated with blind signal separation. Three statistically independent signals s_1, s_2 and s_3 are emitted from unknown locations and arrive by different paths at an array of sensors whose positions are also unknown. It's assumed here, that there are more sensors than signals. Each sensor receives a different unknown mixture of the three signals and so the mixing process may be described by a matrix of complex scalars representing the phase and amplitude values for the different propagation paths. This is referred to as instantaneous mixing. The vector $\mathbf{s}$ in this figure represents the three signal values s_1, s_2 and s_3 at a specific time, the vector $\mathbf{x}$ represents the corresponding sensor outputs and the vector $\mathbf{n}$ represents additive noise which is assumed to be relatively weak. The basic objective in blind signal separation, is to recover a sequence of signal vectors $\mathbf{s}$ from the corresponding sequence of data vectors $\mathbf{x}$ without knowing the mixture matrix $\mathbf{A}$ – quite a challenge, I'm sure you will agree. The entire sequence of signal vectors is represented here by the matrix $\mathbf{S}$ while the corresponding sequence of data vectors is denoted by the matrix $\mathbf{X}$. Each column of these matrices represents a single snapshot in time.

Slide 17 (Principal Components Analysis)

So how can the original signals be recovered? Well the original signals were independent and therefore uncorrelated, but the received signals are correlated as a result of the mixing. The least we should do is decorrelate them – a process which only involves classical second order statistics and standard algorithms from numerical linear algebra. There are several ways to decorrelate the signals but, because it represents the pinnacle of second order adaptive signal processing, I am assuming that we use the method of principal component analysis based on singular value decomposition (or SVD for short). I haven't time to explain this here since it's a standard signal processing technique. Essentially the SVD factorises the data matrix into the form shown on this slide, where $\mathbf{U}$ is a unitary matrix and the matrix $\mathbf{V}$ has orthonormal rows. Note that the matrix $\mathbf{V}$ has more columns than rows. $\mathbf{D}$ is a diagonal matrix and it's assumed that the components contained in the submatrix $\mathbf{D}_s$ are dominant, reflecting the signals that were received – remember there were fewer signals than sensors. Noise alone contibutes to the other rows so we can choose to ignore them and focus on the signal sub-space $\mathbf{V}_s$ – in other words the rows of $\mathbf{V}$ associated with high signal energies. These orthonormal rows satisfy the requirements for sample vectors taken from uncorrelated signals, so in effect we have extracted the uncorrelated signals again. But where has that taken us?

Slide 18 (Hidden Rotation Matrix)

It has, in fact, taken us most of the way, but it hasn't completely solved the problem for the following reason. The matrix $\mathbf{V}_s$ represents uncorrelated signals and has orthonormal rows. However, if you multiply $\mathbf{V}_s$ by an arbitrary unitary matrix $\mathbf{Q}$, you get another matrix $\mathbf{V}_s$-*tilde* with orthonormal rows. In other words, given one set of uncorrelated data, you can generate another set by this type of unitary transformation. If $\mathbf{Q}$ were known, we would be able to deduce the mixture matrix and essentially recover the original signals. But $\mathbf{Q}$ cannot be determined using second order statistics alone so it is often referred to as the hidden rotation matrix. What we can do, however, is exploit the higher order statistics assuming the signals are non-Gaussian.

Slide 19 (Independent Component Analysis)

This figure illustrates the essential features of Independent Component Analysis, in a simple pictorial manner which is also very instructive. Consider three independent non-Gaussian signal waveforms of equal power as shown on the bottom right-hand side of the picture. At each instant in time, the three sample values are treated as the co-ordinates of a point in three-dimensional space and plotted on the three-dimensional figure in the bottom left-hand corner. Repeating this for many time instances produces a three-dimensional graph as shown on the slide. Referred to as a scatter diagram, this type of graph serves to represent the joint probability density function for the three signals. There are three important things to note about the scatter diagram at the bottom left-hand corner. Firstly, it is not spherically symmetric as it would be if the signals were Gaussian distributed. Secondly, the cubic envelope has orthogonal axes reflecting the fact that the signals are uncorrelated. Thirdly, these orthogonal axes correspond to the axes of the graph which means that the signals are not just uncorrelated – they are also statistically independent. This is a much stronger condition which involves the higher order statistics as well as the second order statistics.

When the independent signals were mixed by applying a simple mixing matrix as indicated previously, the scatter diagram for the three mixed signals took the form shown on the top left-hand corner. The envelope of the scatter diagram is now rhomboidal in shape. The axes are no longer orthogonal because the mixed signals are highly correlated.

The mixed signals were then decorrelated and normalised using the PCA technique and this led to the form of scatter diagram shown in the top right-hand corner. The axes of the envelope are orthogonal once again, demonstrating that the signals have been de-correlated. However, these orthogonal axes do not coincide with the axes of the graph revealing that the signals, despite being uncorrelated, are not statistically independent. In order to retrieve the original statistically independent signals, it is necessary to perform a three dimensional rotation on these uncorrelated signals so that the axes are realigned. This is the hidden rotation matrix. It's worth pointing out that if the original signals had been Gaussian distributed, the envelope of this scatter diagram would again be spherically symmetric so we would not have the extra information required to identify the rotation. Any algorithm used to compute the hidden rotation for non-Gaussian signals must exploit the higher order statistics either explicitly or implicitly.

Slide 20 (Higher Order Statistics)

Third order statistics are seldom used since in many applications, the probability distribution for each signal is symmetric and hence the third order cumulants are expected to be zero. It is often sufficient to use fourth order statistics and these form the basis for most of our blind signal separation algorithms. In this slide, I have indicated the essential form of the fourth order cumulants. These involve the average value of the product of four signal samples, each one taken at the same time, from a specific signal as defined by the associated index. Sample x_i is taken from the i^{th} signal and so on. The set of fourth order cumulants defined by letting each of the four indices range from 1 to n, defines the fourth order cumulant tensor. The number of elements in this tensor is obviously given by n^4 where n refers to the number of signals. Now a fundamental property of statistically independent signals, is that the cumulants are equal to zero unless all of the indices are equal. At second order, this means that the correlation matrix is diagonal, in other words, the signals are uncorrelated. At fourth order, it means that the fourth order cumulant tensor should also be diagonal. So for independent component analysis, if we assume that the signals have already been decorrelated, applying the hidden rotation should correspond to diagonalising the fourth order cumulant tensor by means of a unitary transformation. I've tried to illustrate this on the slide, but my graphical skills only rise to three dimensions so the fourth dimension has been left to your imagination.

At first sight, this might look like a straightforward generalisation of the eigenvalue method for matrix diagonalisation. However, the fourth order problem is different and not so well understood. In general, the fourth order cumulant tensor can't be diagonalised exactly in this way, and algorithms must be designed to find the best approximation. It's interesting to note, that some of our most successful algorithms to date involve an iterative sequence of pairwise rotations. Each rotation is aimed at making one pair of signals statistically independent and thus driving the corresponding off-diagonal tensor elements to zero. This is highly reminiscent of the Jacobi algorithm for matrix diagonalisation - but in the case of a fourth order cumulant tensor, these algorithms have not been proven to converge. There is ample scope for some basic mathematical research in this area.

Slide 21 (HF Communications Array)

But that's enough of the theory. I'd now like to give you some examples of blind signal separation in action. The first example relates to HF communications, or shortwave radio. As you are probably aware, shortwave radio is subject to a lot of co-channel interference since this type of radio wave can propagate over very large distances due to reflection from the ionosphere. It is also very difficult to anticipate the strength of a signal or its direction of arrival which vary significantly from time to time. This makes calibration very difficult. We were able to obtain some HF signal data recorded from the type of antenna array shown here and use it to test a blind signal separation algorithm known as BLISS, which was developed by my colleague Ira Clarke. These are some of the results we obtained.

Slide 22 (BLISS Trials Results)

The first audio clip, which you are just about to hear, was obtained in the normal way using a single antenna. It contains a relatively strong signal carrying digital data which accounts for the garbled tones. There is also a much weaker voice signal which is barely audible. The audio clip which you should hear next, is one of the output channels we obtained by applying BLISS to the antenna array data. It is, in fact, the separated voice signal which is now quite audible. I think you will agree that this is a remarkable result given that we had no a-priori knowledge about the signals and were relying entirely on their statistical properties.

Slide 23 (Foetal Heartbeat Analysis)

In QinetiQ we have applied blind signal separation to a wide range of different applications including that of foetal ECG analysis. In this case, an array of ECG sensors is placed on the abdomen of a pregnant woman, the intention being to detect the heartbeat of her unborn baby. However the ECG signal picked up by each sensor is dominated by the mother's own heartbeat.

It would be impractical of course, to calibrate a system like this since the mother is living, breathing and constantly moving. It's an ideal application for blind signal separation and one where we have applied the technique with great success using the QinetiQ BLISS algorithm. Our early results were so encouraging that my colleague, Dr Mark Smith, was funded to develop the idea for clinical application, in collaboration with two of the London teaching hospitals - Queen Charlotte's and Guy's and St Thomas'. This slide shows the prototype system being used in one of many clinical trials. The prototype system works so well, that Mark and his colleagues have been able to separate the ECG signals of twins - and even triplets - from that of their mother and from one another.

Slide 24 (Application to Triplets)

The next slide shows one set of triplet results. The original signals, shown in black, are dominated by the mother's heartbeat and it's difficult to decipher anything else. The signals produced using our blind signal separation equipment are shown in blue and red. The top trace and the third one down come from the mother alone. The three red traces, which have been further processed and displayed on the right hand side, are the three distinct foetal ECG traces. These processed traces are sufficiently clear to reveal potentially useful clinical information concerning the health of each triplet. These results, which have been hailed as an important breakthrough in medical diagnostics, were published recently in "the Lancet".

In another trial, involving a routine single pregnancy, our prototype system produced a very strange trace which warranted further investigation. Independent tests using other medical equipment confirmed that the baby was suffering complete heartblock, and had to be delivered immediately. Mother and baby are now doing very well, and in the opinion of medical experts, blind signal separation may have saved that baby's life.

Slide 25 (Fast-ICA)

For the next few minutes I would like to focus on a relatively new development which I believe is extremely important. An algorithm known as Fast-ICA was published relatively recently by Hyvarinen and Oja from the Helsinki University of Technology. Fast-ICA also uses the fourth order statistics - essentially the kurtosis of each signal - but operates in a slightly different way. I talked earlier about trying to minimise off-diagonal terms in the fourth order cumulant tensor. Well minimising the off-diagonal terms is equivalent to maximising the diagonal terms, so another approach to the ICA problem is to find a weight vector which operates on the received data to produce a signal with maximum kurtosis. This amounts to maximising an expression like the one shown at the top of this slide.

Now this is the sort of objective function that arises in the theory of artificial neural networks. In fact the Fast-ICA algorithm originated from the neural network community whose traditional approach would be to tackle this problem using an adaptive neural network. This would typically involve deriving a stochastic gradient formula, just like the LMS algorithm. However in this case, the weight vector at time $t+1$ is given by the weight vector at time t plus other nonlinear terms. Now I remember claiming at an IEE meeting several years ago, that neural network algorithms would only be useful if one could eliminate the stochastic element – the need to learn the statistics as an integral part of the weight vector computation. Slow convergence often meant having to use the same block of data over and over again and until that problem could be avoided, the neural network approach would be of limited value. The Fast-ICA algorithm has helped to lift the neural network community out of that trap by circumventing that crucial bottleneck.

So what does Fast-ICA do? Hyvarinen and Oja simply noted that we are really interested in the solution to which this update formula ultimately converges. They went on to deduce an equation of the type shown here, which the weight vector must satisfy as *t* tends to infinity. This equation allows for the fact that the data has already been decorrelated and normalised. Now, of course, we don't have a nice linear problem but a nonlinear problem to solve. Nonlinear problems can be tackled by various means and the Fast-ICA algorithm uses an iterative technique as indicated here. It cycles round a loop which involves taking the current weight vector solution, using it to compute the next weight vector solution and then renormalising it. Renormalisation is required, because the weight vector constitutes a single row of the unitary hidden rotation matrix. This algorithm proves to be much faster, since the statistics are computed directly and stochastic learning is not required.

It is worth noting that the last two ICA Conferences were dominated by Fast-ICA as applied to a wide range of problems from financial prediction to image analysis. The neural network community has, at last, been endowed with the type of direct numerical algorithm that we have enjoyed in adaptive signal processing for many years. In a recent study we compared the performance of Fast-ICA and the QinetiQ BLISS algorithm for a range of practical applications. The Fast-ICA algorithm requires less computation and generally produced slightly better signal separation than BLISS. However, it appears to be less reliable and prone to failure on occasions. We now make full use of both algorithms in various circumstances.

Slide 26 (Convolutive Mixing)

I would now like to give you a brief overview of my current research in this area. It concerns the problem of blind signal separation in circumstances where the mixing process is not instantaneous but convolutive in nature. This arises, for example, when the signals arrive, not just via the direct path, but also by other routes due to multi-path propagation etc, as illustrated here for the simple case of two signals and two sensors. This is highly relevant, for example, to acoustic signals in a room as a result of reflections from the walls and other objects, screening of the direct path and so on.

Slide 27 (Channel Model)

In a multi-path situation, the output of each sensor is not just a simple product of the original signal with a complex scalar, but a weighted sum of delayed samples. This corresponds to a convolution of the original signal with the impulse response of the channel which is not generally known. The relationship between the output sequence, the input sequence and the impulse response may be expressed as a simple product of *z*-transforms as shown on this slide. In the special case where *z* lies on the unit circle the *z*-transform is identical to the Fourier transform and the product relationship shown here becomes a statement of the well-known convolution theorem for Fourier transforms.

Slide 28 (Polynomial Matrices)

Using the z-transform, the convolutive mixing process takes the form shown on this slide for the special case of two signals and two sensors. $h_{11}(z)$ denotes the inpulse response for the propagation channel between the first signal and the first sensor and so on. This can obviously be written in a more compact matrix form where elements of the *2x2* matrix $\mathbf{H}(z)$ are now polynomials in z^{-1}. The problem of blind signal separation now involves the identification of an unknown polynomial matrix. This poses a very interesting problem on which Paul Baxter and I have been making significant progress.

Slide 29 (Second Order Stage)

Our approach assumes that the received signals have been strongly decorrelated as indicated on this slide. In other words the signals have been decorrelated, not just instantaneously, but over a wide range of relative delays, denoted by *tau*. This leads to a diagonal cross spectral density matrix of the type shown here. It is further assumed that the individual spectra $sigma_1$ and $sigma_2$ can either be equalised or whitened so that the cross spectral density matrix takes the form *sigma(z)* times the identity matrix.

Slide 30 (Hidden Paraunitary Matrix)

Our method relies on the fact that the hidden rotation matrix which arises in the case of instantaneous mixing and must be unitary, may be generalised in the convolutive case to a hidden polynomial matrix which must be paraunitary. The paraunitary property is defined as shown here in terms of the paraconjugate matrix **H**-*tilde(z)*. This corresponds to taking the Hermitian conjugate of the matrix and also replacing z by $1/z$. For convenience, I have assumed here that **H** is real. I haven't time to go into the significance of this property now. Suffice it to say that our research is devoted to finding ways of identifying the hidden paraunitary matrix using higher order statistics to restore statistical independence. I will be saying more about the algorithm we have developed in a contributed paper to be presented later this morning, so if you are interested in a more detailed description please come along then.

Slide 31 (Future Directions)

Finally, a word on future directions. In general, we use only second order statistics if we have a suitably calibrated array. Otherwise we resort to higher order statistics. But these are extreme cases. In general there may be information - perhaps less accurate information – about the array calibration which can be exploited using second order statistics. But there may also be useful information in the higher order statistics and combing these methods in sensible, computable ways to obtain "semi-blind" algorithms is one of the challenges.

The algorithm which I described for blind signal separation using higher order statistics involved two distinct stages. The first stage used PCA in order to achieve second order independence (decorrelation). The second stage used HOS to identify the hidden rotation matrix. However, in some applications the first stage may need to be corrected or updated so it would be useful to combine the PCA and ICA stages in a suitable manner and this is also the subject of some current research.

There are numerous other challenges and various approaches to addressing them but I think I had better stop here.

Slide 32 (Acknowledgements)

To conclude this talk, I would like to acknowledge the vital contribution made by several of my colleagues from Malvern, from other parts of QinetiQ and also from the University of Leuven. I would also like to thank the UK Ministry of Defence for supporting much of this research. Thank you for your attention.