Synthesis Imaging in Radio Astronomy II
ASP Conference Series, Vol. XXX, 1998
G.B. Taylor, C.L. Carilli, and R. A. Perley (eds.)

10. Self-Calibration

T. Cornwell & E. B. Fomalont
National Radio Astronomy Observatory, Socorro, NM, USA, 87801

Abstract.
In this lecture the principles, techniques, and foibles of self-calibration are discussed.

1. Problems with Ordinary Calibration

Calibrating a synthesis array is one of the most difficult aspects of its operation and, in many cases, is the most important factor in determining the quality of the final deconvolved image. Small quasi-random errors in the amplitude and phase calibration of the visibility data scatter power and so produce an increased level of "rumble" in the weaker regions of the image, and other systematic errors can lead to a variety of artifacts in the image.

The ordinary calibration procedure (see Lecture 5) relies on frequent observations of radio sources of known structure, strength and position in order to determine empirical corrections for time-variable instrumental and environmental factors that cannot be measured, or monitored, directly. The relationship between the visibility $\widetilde{V}_{ij}$ observed at time t on the i-j baseline and the true visibility $V_{ij}(t)$ can be written very generally as

$$\widetilde{V}_{ij}(t) = g_i(t)g_j^*(t)G_{ij}(t)V_{ij}(t) + \varepsilon_{ij}(t) + \epsilon_{ij}(t)\,. \qquad (10\text{-}1)$$

The multiplicative factors $g_i(t)$ and $g_j(t)$ represent the effects of the complex gains of the array elements i and j; $G_{ij}(t)$ is the non-factorable part of the gain on the i-j baseline; $\varepsilon_{ij}(t)$ is an additive offset term; and $\epsilon_{ij}(t)$ is a pure, zero-mean, noise term representing the thermal noise. The effects of $G_{ij}(t)$ and $\varepsilon_{ij}(t)$, which cannot be split into antenna-dependent parts, can usually be reduced to a satisfactory degree by clever design (see Lecture 4), so we will ignore them during this lecture. Equation 10-1 then simplifies to

$$\widetilde{V}_{ij}(t) = g_i(t)g_j^*(t)V_{ij}(t) + \epsilon_{ij}(t)\,. \qquad (10\text{-}2)$$

For simplicity we neglect the effects of time averaging and finite bandwidth, discussed in Lectures 2, 12 and 13; these have relatively little impact here. The *element gain* (usually called the *antenna gain* in radio astronomy) really describes the properties of the elements relative to some reference (usually one array element for phase and a "mean" array element for amplitude). Although this use of the word "gain" may seem confusing, it is quite helpful in lumping all element-based properties together. The gain for any one array element has two contributing components: firstly, a slowly varying instrumental part and secondly, a more rapidly varying part due to the atmosphere (and ionosphere) above the element. Variations in the phase part of the atmospheric component nearly always dominate the overall variation of the element gains (see Sec. 4 of Lecture 5).

Given a calibration source near the region to be imaged, one can solve for the element gains as functions of time. Interpolation of the solutions then provides approximate values for use in correcting the source visibility data. If the equations are overdetermined, then a least-squares technique can be utilized to good effect in overcoming the random errors embodied in the $\epsilon_{ij}(t)$. In particular, for an array in which data from all baselines are correlated and whose elements are identical, when calibrating on a point source of flux density S the variance in the gain estimates due to the receiver noise is

$$\sigma_G^2 = \frac{\sigma_V^2}{(N-3)S^2}, \qquad (10\text{-}3)$$

where σ_V^2 denotes the variance of a visibility datum (assuming all visibilities have equal variance) and N is the number of array elements (Cornwell 1981).

The main drawback to ordinary calibration arises from temporal and spatial variations in the atmosphere (and ionosphere) through which the wavefront passes before reaching the array elements. Values for the $g_i(t)$ inferred from observations of a calibration source may not apply to a source observed at a different time and in a different part of the sky. Hence, the effect of the $g_i(t)$'s cannot be removed completely, and residual errors remain. The level of error varies tremendously with the frequency at which the observations are made and with the lengths of the baselines involved, but on a source of appreciable strength it nearly always overwhelms the error due to the receiver noise term.

Other obstacles to ordinary calibration are the strength (or lack of it) of the calibrators, and any resolved structure they may contain. In some circumstances one may not be able to find a sufficiently strong unresolved calibration source anywhere near the source of interest.

The net effect of this calibration problem depends upon the context. In VLBI, it prevents imaging altogether, whereas for shorter-baseline arrays (such as the VLA and Westerbork) it merely lowers the image quality attainable. Fortunately, progress can be made if the element gains are allowed to be degrees of freedom when determining the sky intensity distribution. *Allowing the element gains to be free parameters is the basic principle of self-calibration.*

2. Redundant Calibration and Self-Calibration

We now discuss the pros and cons of letting the element gains be free parameters. If all baselines are correlated then there are, at any one time, N complex gain errors corrupting the $\frac{1}{2}N(N-1)$ complex visibility measurements. Hence there must be at least $\frac{1}{2}N(N-1) - N$ "good" complex numbers hidden in the data that can be used to constrain the true sky intensity distribution. [1] Let us briefly consider what is lost by using only these "good" numbers. The most obvious losses are the absolute position and strength of the source. The former produces a phase term in the visibility which depends upon the difference in position of the element in an interferometer (see Lecture 1); hence it can be factored out as two element-related quantities. The loss of absolute source strength information

[1] Actually, because absolute phase is meaningless for an interferometer, there are $\frac{1}{2}N(N-1) - (N-1)$ "good" phases and $\frac{1}{2}N(N-1) - N$ "good" amplitudes.

is obvious from Equation 10-2. One also loses the ability to distinguish between various different source structures, but we will show that for large enough numbers of array elements this effect is not too important, since the ratio of the number of constraints to number of degrees of freedom increases.

It is clear what one can expect to lose by letting the element gains be free variables, but the many degrees of freedom embodied in the element gains $g_i(t)$ must still be balanced somehow. There are two different schemes: the explicit use of *redundancy*, and the use of *a priori knowledge* about the object. We shall examine these in turn.

2.1. Redundant calibration

Suppose that the geometry of the interferometer array is designed so that some different pairs of array elements measure the same spacing, or (u, v) sample. As an example, consider a one dimensional linear array of N elements equally spaced, with separation d. All spacings except the longest are measured at least once. In fact there are only $N-1$ different spacings measurable, while there are $\frac{1}{2}N(N-1)$ pairs of elements. This redundancy enables solution for both the $N-1$ true visibility samples, up to a linear phase slope, and the N complex gains, again up to a linear phase slope (Hamaker *et al.* 1977). Since the system of equations is overdetermined, a least-squares method can be employed to good effect in suppressing the effects of receiver noise.

Complete redundancy is not necessary for this approach to work; in fact, since only N complex gains need be solved for, there need be only N redundant spacings. The drawback is that the signal-to-noise ratio of the estimated true visibilities decreases, and nulls can prove disastrous.

Redundant calibration is currently used at the Westerbork Synthesis Radio Telescope.

2.2. Self-calibration

The basis of this approach is that in many cases, even after adding the degrees of freedom in the element gains, the estimation of an adequate model of the brightness is still overdetermined (see Lecture 8). Hence self-calibration is really just another method like 'CLEAN' (Lecture 8, Sec. 2) which is used to interpret the visibility data by introducing some plausible assumptions about the source structure.

Our aim is to produce a model $\widehat{I}$ of the sky intensity distribution, the Fourier transform $\widehat{V}$ of which, when corrected by some complex gain factors, reproduces the observed visibilities to within the noise level. The model $\widehat{I}$ should be astronomically plausible: for example, possible constraints are positivity of brightness and confinement of the structure. (Other, more elaborate, constraints could involve the maximization of some measures of "goodness" of an image; see Lecture 8). One convenient method (Schwab 1980b) of obtaining such agreement is to minimize—by adjusting both the complex element gains g_i and g_j and the model intensity distribution $\widehat{I}$—the sum of squares of residuals

$$\mathcal{S} = \sum_k \sum_{\substack{i,j \\ i \neq j}} w_{ij}(t_k) \left| \widetilde{V}_{ij}(t_k) - g_i(t_k) g_j^*(t_k) \widehat{V}_{ij}(t_k) \right|^2 , \qquad (10\text{-}4)$$

where the $w_{ij}(t_k)$ are weights (purely from signal-to-noise considerations each should be set to the reciprocal of the variance of $\epsilon_{ij}(t_k)$). The time over which the gains should be held constant depends upon the signal-to-noise ratio and upon the variability of the atmosphere (see Sec. 5.3).

An interesting and illuminating connection to ordinary calibration is apparent if Equation 10-4 is re-expressed as

$$\mathcal{S} = \sum_k \sum_{\substack{i,j \\ i \neq j}} w_{ij}(t_k) |\widehat{V}_{ij}(t_k)|^2 \left| X_{ij}(t_k) - g_i(t_k) g_j^*(t_k) \right|^2 , \qquad (10\text{-}5)$$

where

$$X_{ij}(t) = \frac{\widetilde{V}_{ij}(t)}{\widehat{V}_{ij}(t)} . \qquad (10\text{-}6)$$

Division by the model visibilities $\widehat{V}_{ij}(t)$ in effect transforms the object being imaged into a pseudo-point source, though admittedly one with rather strange receiver noise, that can then be used in the ordinary calibration outlined in Section 1.

It is crucial to this gain-solution step that there be too few degrees of freedom (i.e., the element gains $g_i(t)$) to allow the model $\widehat{V}_{ij}(t)$ to be reproduced exactly. If there were, nothing would be achieved. The overdeterminacy also means that errors in the model are averaged down, to an extent dependent on the number of elements in the array. This suggests a possible line of attack in which the model is iteratively refined:

1. Make an initial model of the source using whatever constraints you have on the source structure.

2. Convert the source into a point source using the model.

3. Solve for the complex gains.

4. Find the corrected visibility,

$$V_{ij,\mathrm{corr}}(t) = \frac{\widetilde{V}_{ij}(t)}{g_i(t) g_j^*(t)} . \qquad (10\text{-}7)$$

5. Form a new model from the *corrected* data, again using constraints upon the source structure.

6. Go to (2), unless you are satisfied with the current model.

This approach divides the optimization problem into a part dealing only with the (u, v) data and a part dealing only with the model of the sky brightness. The former can be solved by a simple iterative approach (Schwab 1980b), and in Lecture 8 we learned that both 'CLEAN' (Sec. 2) and the Maximum Entropy Method (MEM, Sec. 4) can be used to solve the latter problem.

Another view of this iterative approach arises from the application of an optimization approach, such as MEM, to gain correction. The unknown gains are added as free variables in the optimization. In the specific case of MEM, the problem is then to choose the image I_k and the gains $g_i(t)$ so as to maximize the image entropy

$$\mathcal{H} = -\sum_k I_k \ln \frac{I_k}{M_k e}\,, \qquad (10\text{-}8)$$

subject to

$$\begin{aligned} \mathcal{S} &= \sum_k \sum_{\substack{i,j \\ i\neq j}} w_{ij}(t_k) \left| \widetilde{V}_{ij}(t_k) - g_i(t_k) g_j^*(t_k) \widehat{V}_{ij}(t_k) \right|^2 \\ &= \text{expected value}\,, \end{aligned} \qquad (10\text{-}9)$$

and

$$\sum_k I_k = \text{estimated value of total flux density}\,, \qquad (10\text{-}10)$$

where $\widehat{V}_{ij}(t)$ is given by the inverse Fourier transform of the MEM image I_k.

The most general approach to solving this optimization problem would vary the image and the gains simultaneously, whereas the iterative approach consists of alternately fixing either the image or the gains, and varying the other as required. The latter is certainly easier to code and seems to work most of the time.

2.3. Redundant calibration or self-calibration?

The relative merits of redundant calibration and of self-calibration are still being debated. The real question is not: *Should redundant calibration be used with an existing array?* (of course it should, if that is possible), but rather: *Should new arrays be designed with redundant spacings?* The main advantage of redundant calibration is that the results are almost model-independent (there is a variable phase shift to worry about), but it is less flexible than self-calibration, and it uses the available signal-to-noise ratio rather less efficiently. A compromise would be to use redundant calibration to get the structure basically correct, and then to use self-calibration to improve the signal-to-noise. In practice, self-calibration is more commonly used simply because many arrays are not instantaneously redundant. Therefore in the rest of this lecture we will concentrate on self-calibration. First, however, we digress slightly to emphasize the links of both schemes with other methods of phase correction.

3. Other Approaches to Phase Correction

The two schemes for phase correction described in Section 2 have two close relatives: the concept of *closure*, and *adaptive optics*.

3.1. Closure quantities

In the early days of radio interferometry, Roger Jennison was faced with the problem of measuring phase information with interferometers which were inherently phase-unstable. He was struck by the fact that an appropriate sum of visibility phases around a closed loop of baselines is free of element-related errors (Jennison 1953, 1958). This can be confirmed by taking the phase part of Equation 10-2,

$$\widetilde{\phi}_{ij}(t) = \phi_{ij}(t) + \theta_i(t) - \theta_j(t) + \text{noise term}\,, \tag{10-11}$$

where $\theta_i(t) = \arg g_i(t)$. Now suppose that a loop of three baselines is formed from elements i, j and k. Then the quantity $\widetilde{C}_{ijk}(t)$, known as the observed *closure phase* [2], is given by

$$\begin{aligned} \widetilde{C}_{ijk}(t) &= \widetilde{\phi}_{ij}(t) + \widetilde{\phi}_{jk}(t) + \widetilde{\phi}_{ki}(t) \\ &= \phi_{ij}(t) + \phi_{jk}(t) + \phi_{ki}(t) + \text{noise term} \\ &= C_{ijk}(t) + \text{noise term}\,. \end{aligned} \tag{10-12}$$

Thus, for an array of three or more elements, and neglecting noise, closure phase is always a good observable. For an array of composed of N elements there are $\frac{1}{2}N(N-1) - (N-1)$ independent closure phases; these are just the "good" constraints mentioned in Section 2.

A *closure amplitude* Γ_{ijkl} can be defined for any loop of 4 elements:

$$\Gamma_{ijkl}(t) = \frac{|\widetilde{V}_{ij}(t)|\,|\widetilde{V}_{kl}(t)|}{|\widetilde{V}_{ik}(t)|\,|\widetilde{V}_{jl}(t)|}\,. \tag{10-13}$$

The amplitudes of the complex gains cancel out of these ratios. Thus, apart from noise, the observed and true closure amplitudes should be identical. There are $\frac{1}{2}N(N-1) - N$ such closure amplitudes.

These closure quantities were of little use until the advent of sufficiently fast computers. Neither closure quantity can be used directly to form an image. However, in the 1970s iterative schemes were developed by Readhead and Wilkinson (1978), Cotton (1979) and others to produce 'CLEAN' images consistent with the closure quantities—see Ekers (1984) for an account of the history of closure phase and self-calibration.

Readhead and Wilkinson (RW) used the following approach to incorporate the closure phases:

1. Make an initial model of the source.

2. For all independent closure phases, use the model to provide estimates of the true phases on two baselines and derive the phase on the other baseline in the loop from the observed closure phase.

3. Form a new model, using 'CLEAN', from the observed visibility amplitudes and the predicted visibility phases.

4. Go to (2), unless you are satisfied with the current model.

---

[2] This terminology is similar to that of closing, or closure, errors in traversed loops, used by surveyors.

Readhead *et al.* (1980) developed a similar algorithm to include the closure amplitudes as constraints. The aspect of choice in part (2) was eliminated in Cotton's (1979) algorithm by utilizing a least-squares technique.

These various approaches have been widely used in VLBI to produce so-called *hybrid images* from the poorly calibrated data that are commonly collected. Only three serious drawbacks are present in the RW–Cotton type algorithms:

1. Proper treatment of noise is difficult because it occurs additively in the vector visibility, not in the amplitude or phase (see Eq. 10-2). Thus it obeys a simple normal distribution in the vector, but a much more-complicated Rice distribution in the phase.

2. For any array with a large number of elements there are very many more possible than independent closure quantities. For a source showing significant structure, the different closure quantities will have varying signal-to-noises, and so in the RW approach it is not easy to choose an optimal set of closure quantities.

3. Calibration effects in radio imaging really do occur in relation to antennas, not baselines, so incorporation of other constraints on, for example, the variability of the atmospheric phase, is simplest in an element-based approach (Cornwell and Wilkinson 1981).

All of these disadvantages are overcome in self-calibration which, since it alters only element gains, must conserve the closure quantities and thus is equivalent to the use of closure quantities (Cornwell and Wilkinson 1981).

3.2. Adaptive optics

Optical "antennas" are typically limited to about one-arcsecond resolution by rapidly varying path length fluctuations due in turn to variations in the refractive index of air (see Woolf 1982 for a good description). One recently developed technique for overcoming this distortion is known as adaptive optics; a well-chosen name since the optics of the element are distorted in order to cancel the effects of the path length variations. A "rubber mirror", which can be distorted at rates up to 1 kHz, is inserted into the light path, and its shape is controlled by a feedback loop designed to optimize the quality of the final image (see, e.g., Muller and Buffington 1974). One of the measures of quality is the sharpness, defined to be the sum of the squares of the pixel values. In an interesting paper, Hamaker *et al.* (1977) show that in redundant spacing interferometry the sharpness is maximized by requiring that all redundant spacings yield the same visibility phase, exactly the same requirement as used in Section 2.1.

The connection between adaptive optics and the scheme outlined in Section 2.2 should be obvious. In both approaches, the phase of the array element is seen as a free variable which can be changed to obtain a plausible image. Fortunately, at radio wavelengths the "fringes" (complex visibilities) can be recorded for each interferometer and the correction can be made subsequently, rather than in real time. Furthermore, since "fringes", rather than the image, can be recorded we can keep track of which pair of elements produced each datum. Dyson (1975) has investigated the latter point in relation to adaptive optics; he has shown that interferometer-based correction requires only one photon per atmospheric coherence time per aperture patch to be corrected, while the image-based correction scheme requires the same rate *per pair* of patches. In the latter the extra photons are lost to decorrelation.

4. Why Does Self-Calibration Work?

No proof of convergence has ever been given for self-calibration, so the exact circumstances under which it works are unknown. Such a proof would be very difficult because of the required use of nonlinear methods of deconvolution such as 'CLEAN' to enforce constraints on the source structure. We do however understand *qualitatively* why it works. There are two, related, reasons:

1. Self-calibration is most successful for arrays with large numbers of elements. The ratio of visibility constraints to unknown gains, $\frac{N-2}{2}$ for phases and $\frac{N(N-3)}{2(N-1)}$ for amplitudes, rises without bound as N increases. Consequently, by allowing the calibration to be a variable only a small amount of information is lost.

2. Sources are relatively simple and can be well represented by a small number of degrees of freedom (in the case of 'CLEAN', the parameters specifying the 'CLEAN' components). Hence the source is, in many cases, effectively oversampled and we can afford to introduce a small number of extra degrees of freedom (the antenna gains). The other side of this is that the (u, v) coverage is usually quite good for the simple sources we are interested in.

The basic requirement is that the total number of degrees of freedom (the number of free gains plus the number of free parameters in the model of the sky brightness distribution), should not be greater than the number of independent visibility measurements (see Lecture 8 for further details).

Self-calibration fails when either the signal-to-noise ratio is too low or the source is too complex (relative to the model). Quantitative estimates of the signal-to-noise requirements can be made; whereas the effect of source complexity is much more difficult to estimate, and further work is needed in this area.

5. Practical Problems in Self-Calibration

We will now consider the details of controlling the self-calibration process. Of all the steps involved in image construction, self-calibration is probably the easiest to perform incorrectly, and so a certain amount of care must be employed when choosing the various parameters. Many of these steps are also described, in more detail, in Lecture 16.

5.1. Specifying the model

In the early days of hybrid imaging great care was taken when producing, usually by model-fitting to the amplitudes, an initial model of the sky brightness; the subsequent convergence depended strongly upon the quality of this model. However, experience with self-calibration algorithms used on data from arrays with relatively modest numbers of elements, such as MERLIN, indicates that for a reasonably simple source, use of an initial point source model may delay but will not prevent convergence—see Cornwell and Wilkinson (1981), for example.

Partially phase-stable arrays such as the VLA usually produce visibility data which, on initial imaging and 'CLEAN'ing, give 'CLEAN' component models which can be used to start self-calibration (even though the associated 'CLEAN' images have only modest dynamic range—typically 10–20 dB).

At any stage in self-calibration *it is important to exclude any features of the model that are due to the very calibration errors we wish to eliminate.* Otherwise, the calibration errors will just be passed through from one iteration to the next. A good rule of thumb when constructing a model from 'CLEAN' components is to exclude all components found after the first negative one. [3] The same rule usually works well in subsequent passes through the self-calibration process. Thus the role that 'CLEAN' or MEM plays in rejecting unsatisfactory models of the sky brightness is apparent; if one used a deconvolution method which did not at least partially reject artifacts due to calibration errors, self-calibration could not increase the dynamic range.

Since the model does not have to be very accurate, an image taken at another frequency will often be useful in speeding convergence. Also, for arrays with many elements, a model made at a higher resolution may be adequate.

5.2. Type of solution and weighting schemes

One can sometimes help convergence by choosing whether to solve Equation 10-4 only for the phases or for both amplitudes and phases. Different weighting schemes can be used to emphasize different parts of the model.

Initially, although the phase errors are usually dominant, the model may represent the true visibility phases very well but the amplitudes very poorly. One such example is the use of a point source model for a symmetrical source such as a Gaussian. Correction of the amplitudes using such a model could produce severe errors in subsequent models. Experience shows that in most cases the quality of the fit of a model to the amplitudes is inferior to the fit to the phases, and so it is often prudent to solve initially for the phase errors only.

The form of the weights can be used to control the solution: in the preferred "natural" weighting scheme, the weights $w_{ij}(t)$ in Equation 10-4 are set to the reciprocal of the expected variance of the errors. The effect of weak visibility points is thus decreased; for visibility functions containing nulls this can be important. If the model has systematic errors then it may be advantageous to make the weights depend upon the (u, v) coordinates. For example, suppose that at high resolution the source is well represented but that an additional amount of extended emission is present. By setting $w_{ij}(t)$ to zero for $\sqrt{u^2 + v^2}$ less than some limit dependent on the source structure we may obtain better estimates for the gain errors than those which would be obtained from all the data.

[3] See Lecture 16 for discussion of possible exceptions to this rule.

5.3. Self-calibration averaging time

Either $\widetilde{V}_{ij}(t)$ or $X_{ij}(t)$ can be averaged over a finite time interval to improve the signal-to-noise ratio. Note that averaging of $X_{ij}(t)$ will not, in general, produce the best signal-to-noise ratio but will correct phase winding that is due to position errors or offsets.

The choice of the optimal averaging time $\tau_{\rm sc}$ obviously depends upon the timescale for gain changes and upon the source strength. The error in the gain estimate due to the receiver noise on a nearly unresolved source is (for good signal-to-noise ratio), for amplitude and phase correction,

$$\sigma_G^2(\tau_{\rm sc}) = \frac{\sigma_V^2(\tau_{\rm sc})}{(N-3)S^2}\,, \tag{10-14}$$

and, for phase correction,

$$\sigma_G^2(\tau_{\rm sc}) = \frac{\sigma_V^2(\tau_{\rm sc})}{(N-2)S^2}\,, \tag{10-15}$$

where S is the approximate flux density of the source, and $\sigma_V^2(\tau)$ is the variance of the receiver noise on each baseline as a function of integration time τ (see Cornwell 1981 for the derivation). One interpretation is that the r.m.s. error in the calculation of the gain of an antenna is approximately the reciprocal of the signal-to-noise ratio for each antenna.

An optimal time between gain solutions can be defined by requiring balance between the errors in the $g_i(t)$ due to gain changes and the errors in the estimates of $g_i(t)$ due to finite signal-to-noise ratio. The condition for self-calibration to be possible is that *the time scale for gain changes should be much greater than the time taken for the noise per antenna to equal the source flux density.*

The errors in the estimated gains must feed back into the image and amplify the noise level. A noise analysis (Cornwell 1981) indicates that on a nearly unresolved source which is sufficiently strong for the errors in the gain estimates to be much less than a radian, the noise level in the background is increased by a small factor $\sqrt{\frac{N-1}{N-3}}$. The corresponding analysis cannot be performed for an extended source, but experience indicates that the noise level is seldom increased by more than a factor of 2 to 3.

5.4. Schwab's ℓ_1 and ℓ_2 solutions

Schwab (1981) has noted that minimization of sums of squares of errors (ℓ_2) is overly sensitive to spuriously discrepant points or outliers. He suggests that instead the ℓ_1 form should be minimized:

$$\mathcal{S} = \sum_k \sum_{\substack{i,j \\ i\neq j}} w_{ij}(t_k) \left| \widetilde{V}_{ij}(t_k) - g_i(t_k) g_j^*(t_k) \widehat{V}_{ij}(t_k) \right| . \tag{10-16}$$

Tests on artificially generated data confirm the superiority of the ℓ_1 minimization algorithm when outliers are present. However, if the noise is normally distributed then the ℓ_2 minimization will, of course, provide superior results. Averaging of the data also alleviates this problem since seriously discrepant points are downweighted in the averages $\langle V_{ij}/\widehat{V}_{ij}\rangle$.

5.5. Spectral line self-calibration

In many spectral line observations the signal-to-noise in a single channel is too poor to allow separate self-calibration of each channel. Instead it is preferable to self-calibrate on the continuum emission and then use the gains so derived to correct the individual channel data. Note that separate bandpass calibration is required (see Lectures 5, 18 and 20).

In cases where different lines appear at different locations, one could form a model having three dimensions, two of space and one of frequency, and then solve the corresponding least-squares problem,

$$\mathcal{S} = \sum_k \sum_l \sum_{\substack{i,j \\ i \neq j}} w_{ij}(t_k, \nu_l) \left| \widetilde{V}_{ij}(t_k, \nu_l) - g_i(t_k) g_j^*(t_k) \widehat{V}_{ij}(t_k, \nu_l) \right|^2 . \qquad (10\text{-}17)$$

5.6. Spurious symmetrization

Suppose that we use a point source model for a slightly resolved source: if the number of array elements is sufficiently small, then the corrected phases will be significantly biased towards zero. As a consequence, after one iteration of self-calibration some features in the image will be seen reflected relative to the point-like component. However, if successive iterations are performed the spurious parts of the image will disappear.

Other, more subtle, symmetrizations are also possible but will disappear if enough iterations are performed. One example has been found by Linfield (1986): in simulations of the VLBA augmented by a high orbit satellite-based antenna, self-calibration failed to correct the gain of the orbiter. His explanation is that since one antenna is at one end of all the long spacings, it is difficult to distinguish between the astronomical structure phase, which is nearly equal on all spacings to the orbiter, and the antenna phase. Thus spurious symmetrization of the fine scale structure occurs. One cure is to calibrate the ground-based spacings internally before introducing the orbiter spacings, and then to allow only the orbiter phase to vary.

5.7. Non-convergence and non-uniqueness

Self-calibration nearly always converges to an answer but, especially for arrays (such as MERLIN) containing small numbers of elements, the final image is non-unique. As should now be apparent, there are a large number of free parameters available to the astronomer: apart from those inherent in the 'CLEAN' algorithm (see Lecture 8) the following can be altered in self-calibration:

1. Number of 'CLEAN' components passed in each iteration.
2. (u, v) range allowed for data to be used in solution.
3. Averaging time.
4. Type of solution and weighting scheme.

However, in most cases, poor choices for these and the 'CLEAN' parameters simply yield an image in which the effect of the corrections is not optimal. Only in cases of exceptionally poor (u, v) coverage (e.g., near declination 0°) and a relatively small number of array elements, ≤ 10, have two, or more, significantly different self-calibrated images been found in practice.

5.8. Baseline-related effects

If the gain errors are not purely element-based then self-calibration will, at some level, fail. The r.m.s. sidelobe level introduced by non-factorable errors is

$$\sigma_{B,C} = \frac{\sigma_{G,C}}{\sqrt{M}}, \tag{10-18}$$

where $\sigma_{G,C}$ is the r.m.s. baseline-related gain error and M is the number of such independent non-factorable errors. For the case of a reasonable synthesis with the VLA, $\sigma_{G,C} = 0.01$ and thus the best VLA image, in the absence of baseline-related calibration, will not have a dynamic range greater than about 35 dB.

Many different effects can lead to non-factorable gain errors. Clark (1981) has enumerated some of these and has described their correctability and relative magnitudes. We shall merely summarize some of these (see Clark's memorandum for further information):

1. Errors due to actual correlator problems. These are very unlikely in a digital correlator. They may be correctable if they are sufficiently constant with time.
2. Bandpass mismatches. These do not factor out on an antenna basis. They can, in principle, be corrected if the individual bandpasses are known. They are exacerbated by poorly adjusted delays.
3. Random, varying pointing errors. Simple self-calibration cannot correct for these if the size of the emission region is comparable to the main lobe of the primary beam $A(l, m)$ of the array elements.
4. Non-isoplanaticity of the atmosphere, i.e., different parts of the field of view to be imaged are seen through different cells in the atmosphere. Schwab (1984b) has outlined a possible solution to this problem, which, owing to its complexity, has never been tried out; similar, but less unwieldy algorithms (C. R. Subrahmanya 1988, in preparation) are under development for use in conjunction with the Giant Metrewave Radio Telescope (GMRT), now under construction in India.
5. Finite integration time and/or bandwidth. The latter can, in principle, be corrected, but this may be difficult to do in practice.
6. Incorrectly set sampling levels in the quantizers preceding the correlator.
7. Faulty analog quadrature networks.

All of these effects, save the first, are minimized by locating the source at the phase tracking center. The calibration and correction of baseline-based effects is discussed in Lecture 16.

6. Bibliography

A good and extensive review article on self-calibration appears in the 1984 Edition of the *Annual Review of Astronomy and Astrophysics* (Pearson and Readhead 1984).

References

Clark, B. G. 1981, "Orders of magnitude of some instrumental effects", VLA Scientific Memorandum No. 137, NRAO.

Cornwell, T. J. 1981, "An error analysis of calibration", VLA Scientific Memorandum No. 135, NRAO.

Cornwell, T. J. and Wilkinson, P. N. 1981, "A new method for making radio maps with unstable radio interferometers", *Mon. Not. Roy. Astron. Soc.*, **196**, 1067–1086.

Cotton, W. D. 1979, "A method of measuring compact structure in radio sources using VLBI observations", *Astron. J.*, **84**, 1122–1128.

Dyson, F. J. 1975, "Photon noise and atmospheric noise in active optical systems", *J. Opt. Soc. Am.*, **65**, 551–558.

Ekers, R. D. 1984, "The almost serendipitous discovery of self-calibration", in *Serendipitous Discoveries in Radio Astronomy*, Proceedings of a Workshop Held at the NRAO on May 4, 5, 6, 1983, K. I. Kellermann and B. Sheets, Eds., NRAO (Green Bank, WV), pp. 154–159.

Hamaker, J. P., O'Sullivan, J. D., and Noordam, J. E. 1977, "Image sharpness, Fourier optics, and redundant spacing interferometry", *J. Opt. Soc. Am.*, **67**, 1122–1123.

Jennison, R. C. 1953, *The Measurement of the Fine Structure of the Cosmic Radio Sources*, Ph. D. Thesis, University of Manchester.

Jennison, R. C. 1958, "A phase sensitive interferometer technique for the measurement of the Fourier transforms of spatial brightness distributions of small angular extent", *Mon. Not. Roy. Astron. Soc.*, **118**, 276–284.

Linfield, R. P. 1986, "A modified hybrid mapping technique for VLBI data", *Astron. J.*, **92**, 213–218.

Muller, R. A. and Buffington, A. 1974, "Real-time correction of atmospherically degraded telescope images through image sharpening", *J. Opt. Soc. Am.*, **64**, 1200–1210.

Pearson, T. J. and Readhead, A. C. S. 1984, "Image-formation by self-calibration in radio astronomy", *Ann. Rev. Astron. Astrophys.*, **22**, 97–130.

Readhead, A. C. S., Walker, R. C., Pearson, T. J., and Cohen, M. H. 1980, "Mapping radio sources with uncalibrated visibility data", *Nature*, **285**, 137–140.

Readhead, A. C. S. and Wilkinson, P. N. 1978, "The mapping of compact sources from VLBI data", *Astrophys. J.*, **223**, 25–36.

Schwab, F. R. 1980, "Adaptive calibration of radio interferometer data", *Proc. S.P.I.E.*, **231**, 18–25.

Schwab, F. R. 1981, "Robust solutions for antenna gains", VLA Scientific Memorandum No. 136, NRAO.

Schwab, F. R. 1984, "Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry", *Astron. J.*, **89**, 1076–1081.

Woolf, N. J. 1982, "High resolution imaging from the ground", *Ann. Rev. Astron. Astrophys.*, **20**, 367–398.