Nonstationarity: patching

Jon Claerbout

There are many reasons for cutting data planes or image planes into overlapping pieces (patches), operating on the pieces, and then putting them back together again, as depicted in Figure 1. The earth's dip varies with lateral location and depth. The dip spectrum and spatial spectrum thus also varies. The dip itself is the essence of almost all earth mapping, and its spectrum plays an important role in the estimation any earth properties. In statistical estimation theory, the word to describe changing statistical properties is "**nonstationary**".

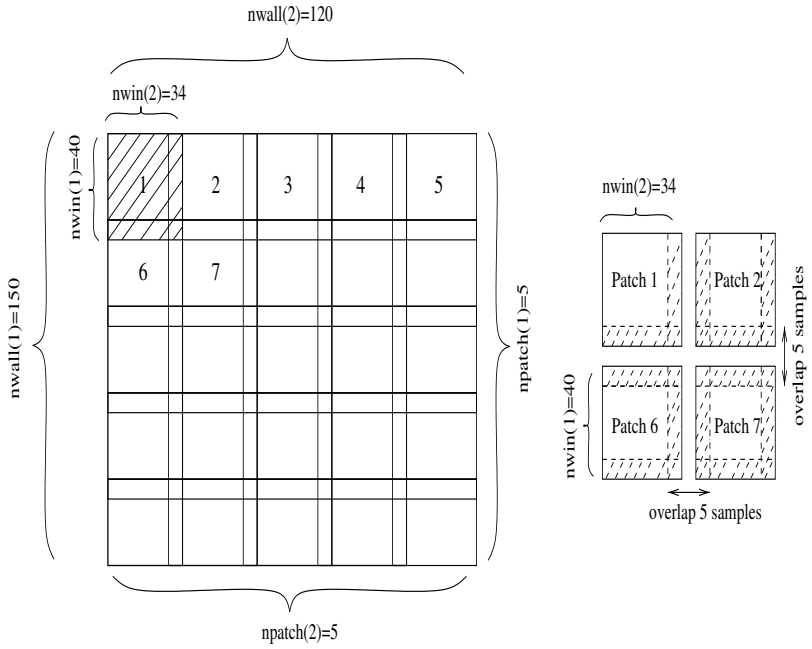

Figure 1: Decomposing a wall of information into windows (also called patches). Left is an example of a 2-D space input to module `patch`. Right shows a close-up of the output (top left corner).

We begin this chapter with basic patching concepts along with supporting utility code. The language of this chapter, *patch, overlap, window, wall,* is two-dimensional, but it may as well be three-dimensional, *cube, subcube, brick,* or one-dimensional, *line, interval.* We sometimes use the language of windows on a wall. But since we usually want to have overlapping windows, better imagery would be to say we assemble a quilt from patches.

The codes are designed to work in any number of dimensions. After developing the infrastructure, we examine some two-dimensional, time- and space-variable applications: adaptive steep-dip rejection, noise reduction by prediction, and segregation of signals and noises.

PATCHING TECHNOLOGY

A plane of information, either data or an image, say `wall(nwall1, nwall2)`, will be divided up into an array of overlapping **window**s each window of size `(nwind1,nwind2)`. To choose the number of windows, you specify `(npatch1,npatch2)`. Overlap on the 2-axis is measured by the fraction `(nwind2*npatch2)/nwall2`. We turn to the language of F90 which allows us to discuss N-dimensional hypercubes almost as easily as two-dimensional spaces. We define an N-dimensional volume (like the wall) with the vector `nwall= (nwall1, nwall2, ...)`. We define subvolume size (like a 2-D window) with the vector `nwind=(nwind1, nwind2, ...)`. The number of subvolumes on each axis is `npatch=(npatch1, npatch2, ...)`. The operator `patch` on the following page simply grabs one patch from the wall, or when used in adjoint form, it puts the patch back on the wall. The number of patches on the wall is `product(npatch)`. Getting and putting all the patches is shown later in module `patching` on page 6.

The i-th patch is denoted by the scalar counter `ipatch`. Typical patch extraction begins by taking `ipatch`, a C linear index, and converting it to a multidimensional subscript `jj` each component of which is less than `npatch`. The patches cover all edges and corners of the given data plane (actually the hypervolume) even where `nwall/npatch` is not an integer, even for axes whose length is not an integer number of the patch length. Where there are noninteger ratios, the spacing of patches is slightly uneven, but we'll see later that it is easy to reassemble seamlessly the full plane from the patches, so the unevenness does not matter. You might wish to review the utilities `line2cart` and `cart2line` on page ?? which convert between multidimensional array subscripts and the linear memory subscript before looking at the patch extraction-putback code: The cartesian vector `jj` points to the beginning of a patch, where on the wall the (1,1,..) coordinate of the patch lies. Obviously this begins at the beginning edge of the wall. Then we pick `jj` so that the last patch on any axis has its last point exactly abutting the end of the axis. The formula for doing this would divide by zero for a wall with only one patch on it. This case arises legitimately where an axis has length one. Thus we handle the case `npatch=1` by abutting the patch to the beginning of the wall and forgetting about its end. As in any code mixing integers with floats, to guard against having a floating-point number, say 99.9999, rounding down to 99 instead of up to 100, the rule is to always add .5 to a floating point number the moment before converting it to an integer. Now we are ready to sweep a window to or from the wall. The number of points in a window is `size(wind)` or equivalently `product(nwind)`.

Figure 2 shows an example with five nonoverlapping patches on the 1-axis and many overlapping patches on the 2-axis.

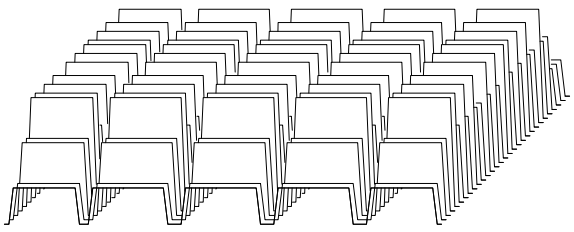

Figure 2: A plane of identical values after patches have been cut and then added back. Results are shown for `nwall=(100,30)`, `nwind=(17,6)`, `npatch=(5,11)`. For these parameters, there is gapping on the horizontal axis and overlap on the depth axis.

user/gee/patch.c

```
void patch_lop (bool adj, bool add, int nx, int ny,
                float* wall, float* wind)
/*< patch operator >*/
{
    int i, j, shift;

    sf_adjnull (adj, add, nx, ny, wall, wind);
    sf_line2cart(dim, npatch, ipatch, jj);
    for(i = 0; i < dim; i++) {
        if(npatch[i] == 1) {
            jj[i] = 0;
        } else if (jj[i] == npatch[i]-1) {
            jj[i] = nwall[i] - nwind[i];
        } else {
            jj[i] = jj[i]*(nwall[i] - nwind[i])/(npatch[i] - 1.0);
        }
    }

    /* shift till the patch start */
    shift = sf_cart2line(dim, nwall, jj);
    for(i = 0; i < ny; i++) {
        sf_line2cart(dim, nwind, i, ii);
        j = sf_cart2line(dim, nwall, ii) + shift;
        if (adj) wall[j] += wind[i];
        else     wind[i] += wall[j];
    }
}
```

Weighting and reconstructing

The adjoint of extracting all the patches is adding them back. Because of the overlaps, the adjoint is not the inverse. In many applications, **inverse patching** is required; i.e. patching things back together seamlessly. This can be done with weighting functions. You can have any weighting function you wish and I will provide you the patching reconstruction operator $\tilde{\mathbf{I}}_p$ in

$$\tilde{\mathbf{d}} \quad = \quad [\mathbf{W}_{\text{wall}}\mathbf{P}^{\text{T}}\mathbf{W}_{\text{wind}}\mathbf{P}]\mathbf{d} \quad = \quad \tilde{\mathbf{I}}_p\,\mathbf{d} \tag{1}$$

where $\mathbf{d}$ is your initial data, $\tilde{\mathbf{d}}$ is the reconstructed data, $\mathbf{P}$ is the patching operator, $\mathbf{P}^{\text{T}}$ is adjoint patching (adding the patches). $\mathbf{W}_{\text{wind}}$ is your chosen weighting function in the window, and $\mathbf{W}_{\text{wall}}$ is the weighting function for the whole wall. You specify any $\mathbf{W}_{\text{wind}}$ you like, and module `mkwallwt` below builds the weighting function $\mathbf{W}_{\text{wall}}$ that you need to apply to your wall of reconstructed data, so it will undo the nasty effects of the overlap of windows and the shape of your window-weighting function. You do not need to change your window weighting function when you increase or decrease the amount of overlap between windows because $\mathbf{W}_{\text{wall}}$ takes care of it. The method is to use adjoint `patch` on the previous page to add the weights of each window onto the wall and finally to invert the sum wherever it is non-zero. (You lose data wherever the sum is zero).

No matrices are needed to show that this method succeeds, because data values are never mixed with one another. An equation for any reconstructed data value $\tilde{d}$ as a function of the original value d and the weights w_i that hit d is $\tilde{d} = (\sum_i w_i d)/\sum_i w_i = d$. Thus, our process is simply a "partition of unity."

To demonstrate the program, I made a random weighting function to use in each window with positive random numbers. The general strategy allows us to use different weights in different windows. That flexibility adds clutter, however, so here we simply use the same weighting function in each window.

The operator $\tilde{\mathbf{I}}_p$ is called "**idempotent**." The word "idempotent" means "self-power," because for any N, $0^N = 0$ and $1^N = 1$, thus the numbers 0 and 1 share the property that raised to any power they remain themselves. Likewise, the patching reconstruction operator multiplies every data value by either one or zero. Figure 3 shows the result obtained when a plane of identical constant values $\mathbf{d}$ is passed into the patching reconstruction operator $\tilde{\mathbf{I}}_p$. The result is constant on the 2-axis, which confirms that there is adequate sampling on the 2-axis, and although the weighting function is made of random numbers, all trace of random numbers has disappeared from the output. On the 1-axis the output is constant, except for being zero in gaps, because the windows do not overlap on the 1-axis.

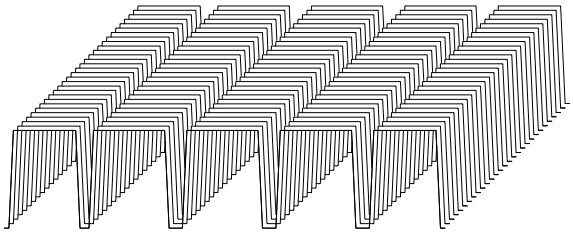

Figure 3: A plane of identical values passed through the idempotent patching reconstruction operator. Results are shown for the same parameters as Figure 2.

Module `patching` assists in reusing the patching technique. It takes a linear operator $\mathbf{F}$. as its argument and applies it in patches. Mathematically, this is $[\mathbf{W}_{\text{wall}}\mathbf{P}^{\text{T}}\mathbf{W}_{\text{wind}}\mathbf{F}\mathbf{P}]\mathbf{d}$. It is assumed that the input and output sizes for the operator `oper` are equal.

user/gee/mkwallwt.c

```
void mkwallwt(int dim      /* number of dimensions */,
              int* npatch /* number of patches [dim] */,
              int* nwall  /* data size [dim] */,
              int* nwind  /* patch size [dim] */,
              float* windwt /* window weighting (input) */,
              float* wallwt /* wall weighting (output) */)
/*< make wall weight >*/
{
    int i, j, ip, np, n, nw;

    np = 1;
    n = 1;
    nw = 1;

    for (j=0; j < dim; j++) {
        np *= npatch[j];
        n *= nwall[j];
        nw *= nwind[j];
    }

    for (i = 0; i < n; i++) {
        wallwt[i] = 0.;
    }

    patch_init(dim, npatch, nwall, nwind);

    for (ip=0; ip < np; ip++) {
        patch_lop(true, true, n, nw, wallwt, windwt);
        patch_close ();
    }

    for (i = 0; i < n; i++) {
        if ( wallwt[i] != 0.) wallwt[i] = 1. / wallwt[i];
    }
}
```

user/gee/patching.c

```
1     for (i=0; i < n; i++) data[i] = 0.;
2
3     patch_init(dim, npatch, nwall, nwind);
4     for (ip = 0; ip < np; ip++) {
5         /* modl -> winmodl */
6         patch_lop(false, false, n, nw, modl, winmodl);
7         /* winmodl -> windata */
8         oper(false, false, nw, nw, winmodl, windata);
9         /* apply window weighting */
10        for (iw=0; iw < nw; iw++) windata[iw] *= windwt[iw];
11        /* data <- windata */
12        patch_lop(true, true, n, nw, data, windata);
13        patch_close();
14    }
15
16    /* windwt -> wallwt */
17    mkwallwt(dim, npatch, nwall, nwind, windwt, wallwt);
18
19    /* apply wall weighting */
20    for (i=0; i < n; i++) data[i] *= wallwt[i];
```

2-D filtering in patches

A way to do time- and space-variable filtering is to do invariant filtering within each patch. Typically, we apply a filter, say $\mathbf{F}_p$, in each patch. The composite operator, filtering in patches, $\tilde{\mathbf{F}}$, is given by

$$\tilde{\mathbf{d}} \quad = \quad [\mathbf{W}_{\text{wall}}\mathbf{P}^{\mathrm{T}}\mathbf{W}_{\text{wind}}\mathbf{F}_p\mathbf{P}]\ \mathbf{d} \quad = \quad \tilde{\mathbf{F}}\ \mathbf{d} \tag{2}$$

I built a triangular weighting routine `tentn()` that tapers from the center of the patch of the filter's *outputs* towards the edges. Accomplishing this weighting is complicated by (1) the constraint that the filter must not move off the edge of the input patch and (2) the alignment of the input and the output. The layout for prediction-error filters is shown in Figure 4. We need a weighting function that vanishes where the filter has no outputs.

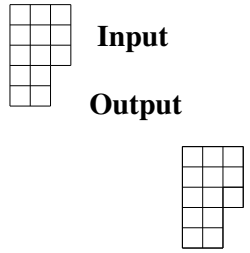

Figure 4: Domain of inputs and outputs of a two-dimensional prediction-error filter.

The amplitude of the weighting function is not very important because we have learned how to put signals back together properly for arbitrary weighting functions. We can use any

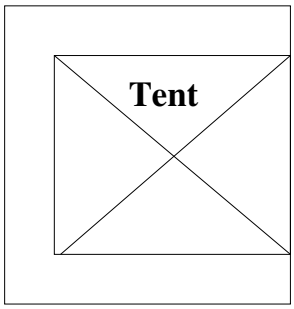

Figure 5: Placement of tent-like weighting function in the space of filter inputs and outputs.

pyramidal or tent-like shape that drops to zero outside the domain of the filter output. The job is done by subroutine `tentn()`. A new parameter needed by `tentn` is `a`, the coordinate of the beginning of the tent.

user/gee/tent.c

```
    /* loop in the window */
    for (i=0; i < nw; i++) {
        sf_line2cart(dim, nwind, i, x);

        windwt[i] = 1.;
        for (j=0; j < dim; j++) {
            if (x[j] >= start[j] && x[j] <= end[j]) {
                w = (x[j]-mid[j])/wid[j];
                windwt[i] *= SF_MAX(0.,1.-fabs(w));
            } else {
                windwt[i] = 0.;
            }
        }
    }
```

In applications where triangle weights are needed on the *inputs* (or where we can work on a patch without having interference with edges), we can get "triangle tent" weights from `tentn()` if we set filter dimensions and lags to unity, as shown in Figure 6.

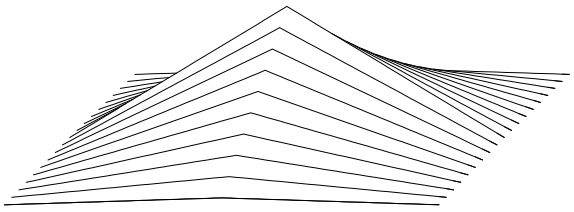

Figure 6: Window weights from `tentn()` with `nwind=(61,19), center=(31,1), a=(1,1)` .

Triangle weighting functions can sum to a constant if the spacing is such that the midpoint of one triangle is at the beginning of the next. I imagined in two dimensions that something similar would happen with shapes like Egyptian pyramids of Cheops, $2 - |x - y| + |x + y|$. Instead, the equation $(1 - |x|)(1 - |y|)$ which has the tent-like shape shown in Figure 6 adds up to the constant flat top shown in Figure 7. (To add interest to Figure 7,

I separated the windows by a little more than the precise matching distance.) In practice we may chose window shapes and overlaps for reasons other than the constancy of the sum of weights, because `mkwallwt` on page 5 accounts for that.

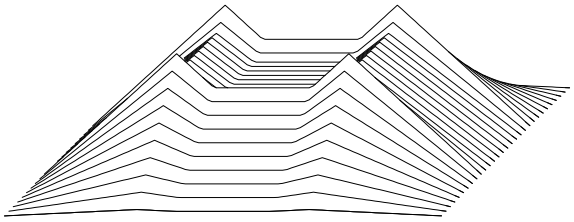

Figure 7: (Inverse) wall weights with `n1=100, w1=61, k1=2, n2=30, w2=19, k2=2`

Finally is an example of filtering a plane of uniform constants with an impulse function. The impulse function is surrounded by zeros, so the filter output patches are smaller than the input patches back in Figure 3. Here in Figure 8, both axes need more window density.

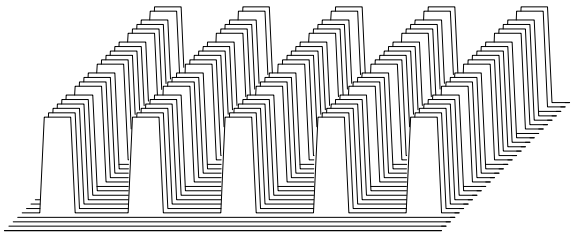

Figure 8: Filtering in patches Mid with the same parameters as in Figures 2 and 3. Additionally, the filter parameters are `a1=11 a2=5 lag1=6 lag2=1` . Thus, windows are centered on the 1-axis and pushed back out the 2-axis.

Designing a separate filter for each patch

Recall the prediction-error filter subroutine `find_pef()` on page ??. Given a data plane, this subroutine finds a filter that tends to whiten the spectrum of that data plane. The output is white residual. Now suppose we have a data plane where the dip spectrum is changing from place to place. Here it is natural to apply subroutine `find_pef()` in local patches. This is done by subroutine `find_lopef()`. The output of this subroutine is an array of helix-type filters, which can be used, for example, in a local convolution operator `loconvol` We notice that when a patch has fewer regression equations than the filter has coefficients, then the filter is taken to be that of the previous patch.

Triangular patches

I have been running patching code for several years and my first general comment is that realistic applications often call for patches of different sizes and different shapes. (Tutorial, non-interactive C code is poorly suited to this need.) Raw seismic data in particular seems more suited to triangular shapes. It is worth noting that the basic concepts in this chapter have ready extension to other shapes. For example, a rectangular shape could be duplicated into two identical patches; then data in one could be zeroed above the diagonal and in the other below; you would have to allow, of course, for overlap the size of the filter. Module `pef` on page ?? automatically ignores the zeroed portion of the triangle, and it is irrelevant

user/gee/lopef.c

```
    patch_init(dim, npatch, nwall, nwind);
    for (ip=0; ip < np; ip++) {
        bb = aa+ip;

        patch_lop(false, false, n, nw, wall, windata);
        if (NULL != mask) {
            patch_lop(false, false, n, nw, mask, winmask);
            for (iw=0; iw < nw; iw++) {
                known[iw] = (winmask[iw] != 0.);
            }
            find_mask(nw, known, bb);
        }
        for (mis=iw=0; iw < nw; iw++) {
            if (!bb->mis[iw]) mis++;
        }
        if (mis > nh) { /* enough equations */
            find_pef(nw, windata, bb, nh);
        } else if (ip > 1) { /* use last PEF */
            for (ih=0; ih < nh; ih++) {
                bb->flt[ih] = (bb-1)->flt[ih];
            }
        }
        patch_close();
    }
```

user/gee/loconvol.c

```
void loconvol_init(sf_filter aa_in)
/*< initialize with the first filter >*/
{
    aa = aa_in;
}

void loconvol_lop(bool adj, bool add, int nx, int ny,
                  float *xx, float *yy)
/*< convolve >*/
{
    sf_helicon_init(aa);
    aa++;

    sf_helicon_lop(adj, add, nx, ny, xx, yy);
}
```

what `mis2()` on page ?? does with a zeroed portion of data, if a triangular footprint of weights is designed to ignore its output.

EXERCISES:

1 Code the linear operator $\mathbf{W}_{\text{wall}}\mathbf{P}^{\text{T}}\mathbf{W}_{\text{wind}}\mathbf{P}$ including its adjoint.

2 **Smoothing program.** Some familiar operations can be seen in a new light when done in patches. Patch the data. In each patch, find the mean value. Replace each value by the mean value. Reconstruct the wall.

3 **Smoothing while filling missing data.** This is like smoothing, but you set window weights to zero where there is no data. Because there will be a different set of weights in each window, you will need to make a simple generalization to `mkwallwt` on page 5.

4 **Gain control.** Divide the data into patches. Compute the square root of the sum of the squares of all the data in each patch. Divide all values in that patch by this amount. Reassemble patches.

STEEP-DIP DECON

Normally, when an autoregression filter (PEF) predicts a value at a point it uses values at earlier points. In practice, a gap may also be set between the predicted value and the earlier values. What is not normally done is to supplement the fitting signals on nearby traces. That is what we do here. We allow the prediction of a signal to include nearby signals at earlier times. The times accepted in the goal are inside a triangle of velocity less than about the water velocity. The new information allowed in the prediction is extremely valuable for water-velocity events. Wavefronts are especially predictable when we can view them along the wavefront (compared to perpendicular or at some other angle from the wavefront). It is even better on land, where noises move more slowly at irregular velocities, and are more likely to be aliased.

Using `lopef` on the previous page, the overall process proceeds independently in each of many overlapping windows. The most important practical aspect is the filter masks, described next.

Dip rejecting known-velocity waves

Consider the two-dimensional filter

$$\begin{array}{ccc} & +1 & \\ -1 & 0 & -1 \\ & +1 & \end{array} \tag{3}$$

When this this filter is applied to a field profile with 4 ms time sampling and 6 m trace

spacing, it should perfectly extinguish 1.5 km/s water-velocity noises. Likewise, the filter

$$\begin{array}{ccccc} & & +1 & & \\ & & 0 & & \\ & & 0 & & \\ -1 & & 0 & & -1 \\ & & 0 & & \\ & & 0 & & \\ & & +1 & & \end{array} \tag{4}$$

should perfectly extinguish water noise when the trace spacing is 18 m. Such noise is, of course, spatially aliased for all temporal frequencies above 1/3 of Nyquist, but that does not matter. The filter extinguishes them perfectly anyway. Inevitably, the filter cannot both extinguish the noise and leave the signal untouched where the alias of one is equal to the other. So we expect the signal to be altered where it matches aliased noise. This simple filter does worse than that. On horizontal layers, for example, signal wavelets become filtered by $(1, 0, 0, -2, 0, 0, 1)$. If the noise is overwhelming, this signal distortion is a small price to pay for eliminating it. If the noise is tiny, however, the distortion is unforgivable. In the real world, data-adaptive deconvolution is usually a good compromise.

The two-dimensional deconvolutions filters we explore here look like this:

$$\begin{array}{ccccccccc} x & x & x & x & x & x & x & x & x \\ x & x & x & x & x & x & x & x & x \\ . & x & x & x & x & x & x & x & . \\ . & x & x & x & x & x & x & x & . \\ . & x & x & x & x & x & x & x & . \\ . & . & x & x & x & x & x & . & . \\ . & . & x & x & x & x & x & . & . \\ . & . & x & x & x & x & x & . & . \\ . & . & . & x & x & x & . & . & . \\ . & . & . & x & x & x & . & . & . \\ . & . & . & x & x & x & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & 1 & . & . & . & . \end{array} \tag{5}$$

where each . denotes a zero and each x denotes a (different) adjustable filter coefficient that is chosen to minimize the power out.

You can easily imagine variations on this shape, such as a diamond instead of a triangle. I invite you to experiment with the various shapes that suggest themselves.

Tests of steep-dip decon on field data

Low-velocity noises on shot records are often not fully suppressed by stacking because the noises are spatially aliased. Routine field arrays are not perfect and the noise is often

extremely strong. An interesting, recently-arrived data set worth testing is shown in Figure 9.

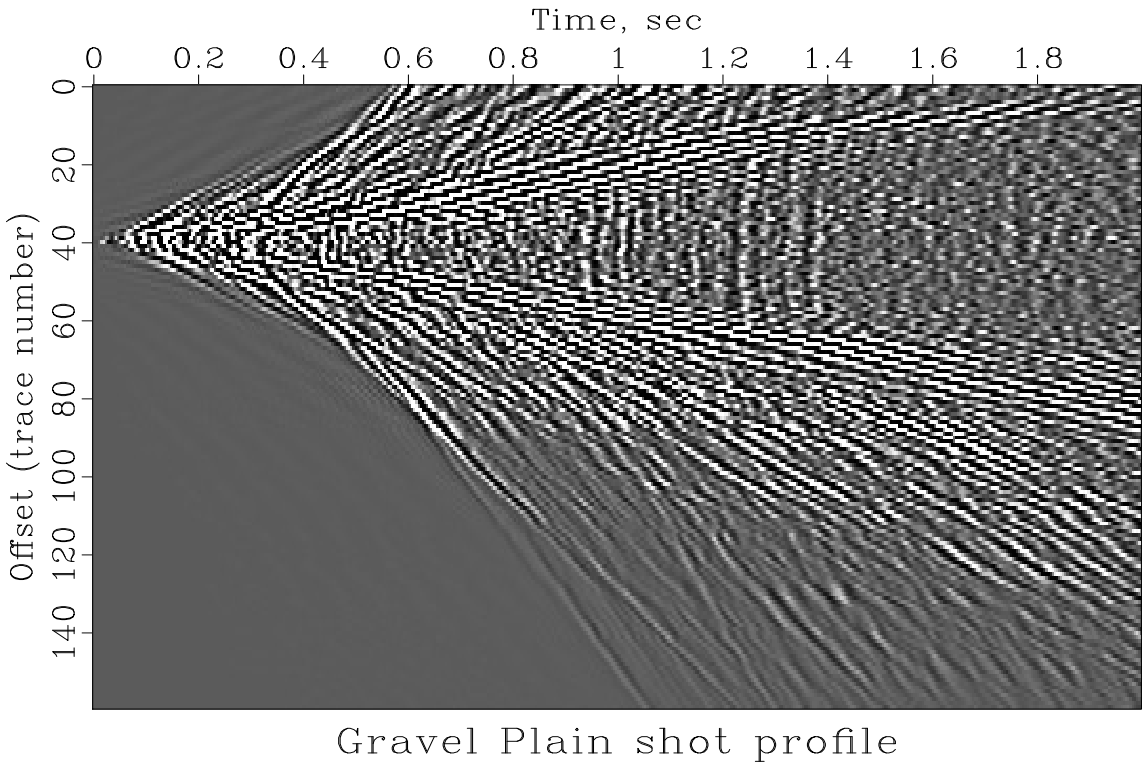

Figure 9: Gravel plain ground roll (Middle East) Worth testing.

I scanned the forty **Yilmaz** and **Cumro** shot profiles for strong low-velocity noises and I selected six examples. To each I applied an AGC that is a slow function of time and space (triangle smoothing windows with triangle half-widths of 200 time points and 4 channels). Because my process simultaneously does both low-velocity rejection and deconvolution, I prepared more traditional 1-D deconvolutions for comparison. This is done in windows of 250 time points and 25 channels, the same filter being used for each of the 25 channels in the window. In practice, of course, considerably more thought would be given to optimal window sizes as a function of the regional nature of the data. The windows were overlapped by about 50%. The same windows are used on the steep-dip deconvolution.

It turned out to be much easier than expected and on the first try I got good results on all all six field profiles tested. I have not yet tweaked the many adjustable parameters. As you inspect these deconvolved profiles from different areas of the world with different recording methods, land and marine, think about how the stacks should be improved by the deconvolution. Stanford Exploration Project report 77 (SEP-77) shows the full suite of results. Figure 10 is a sample of them.

Unexpectedly, results showed that 1-D deconvolution also suppresses low-velocity noises. An explanation can be that these noises are often either low-frequency or quasimonochromatic.

As a minor matter, fundamentally, my code cannot work ideally along the side bound-

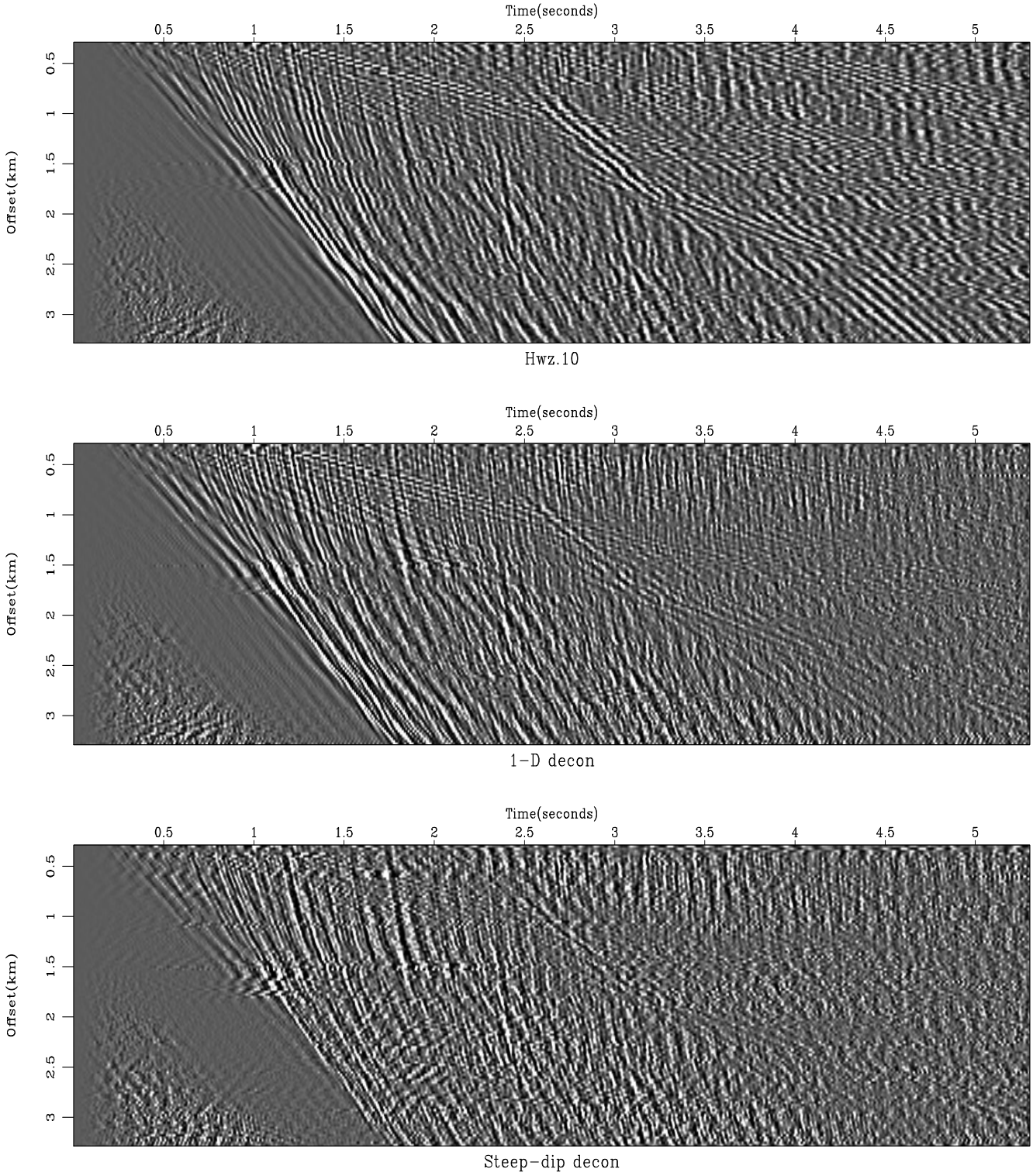

Figure 10: Top is a North African vibrator shot profile (Y&C #10) after AGC. Middle is gapped 1-D decon. Bottom is steep-dip decon.

aries because there is no output (so I replaced it by the variance scaled input). With a little extra coding, better outputs could be produced along the sides if we used spatially one-sided filters like

$$\begin{array}{ccccc} x & x & x & x & x \\ . & x & x & x & x \\ . & x & x & x & x \\ . & . & x & x & x \\ . & . & x & x & x \\ . & . & . & x & x \\ . & . & . & x & x \\ . & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & 1 \end{array} \tag{6}$$

These would be applied on one side of the shot and the opposite orientation would be applied on the other side. With many kinds of data sets, such as off-end marine recording in which a ship tows a hydrophone streamer, the above filter might be better in the interior too.

Are field arrays really needed?

Field arrays cancel random noise but their main function, I believe, is to cancel low-velocity coherent noises, something we now see is handled effectively by steep-dip deconvolution. While I do not advocate abandoning field arrays, it is pleasing to notice that with the arrival of steep-dip deconvolution, we are no longer so dependent on field arrays and perhaps coherent noises can be controlled where field arrays are impractical, as in certain 3-D geometries. A recently arrived 3-D shot profile from the sand dunes in the Middle East is Figure 11. The strong hyperbolas are **ground roll** seen in a line that does not include the shot. The open question here is, how should we formulate the problem of ground-roll removal in 3-D?

Which coefficients are really needed?

Steep-dip decon is a heavy consumer of computer time. Many small optimizations could be done, but more importantly, I feel there are some deeper issues that warrant further investigation. The first question is, how many filter coefficients should there be and where should they be? We would like to keep the number of nonzero filter coefficients to a minimum because it would speed the computation, but more importantly I fear the filter output might be defective in some insidious way (perhaps missing primaries) when too many filter coefficients are used. Perhaps if 1-D decon were done sequentially with steep-dip decon the number of free parameters (and hence the amount of computer time) could be dropped even further. I looked at some of the filters and they scatter wildly with the Nyquist frequency (particularly those coefficients on the trace with the "1" constraint). This suggests using a damping term on the filter coefficients, after which perhaps the magnitude of a filter coefficient will be a better measure of whether this practice is really helpful. Also, it would, of course, be fun to get some complete data sets (rather than a single shot profile) to see the difference in the final stack.

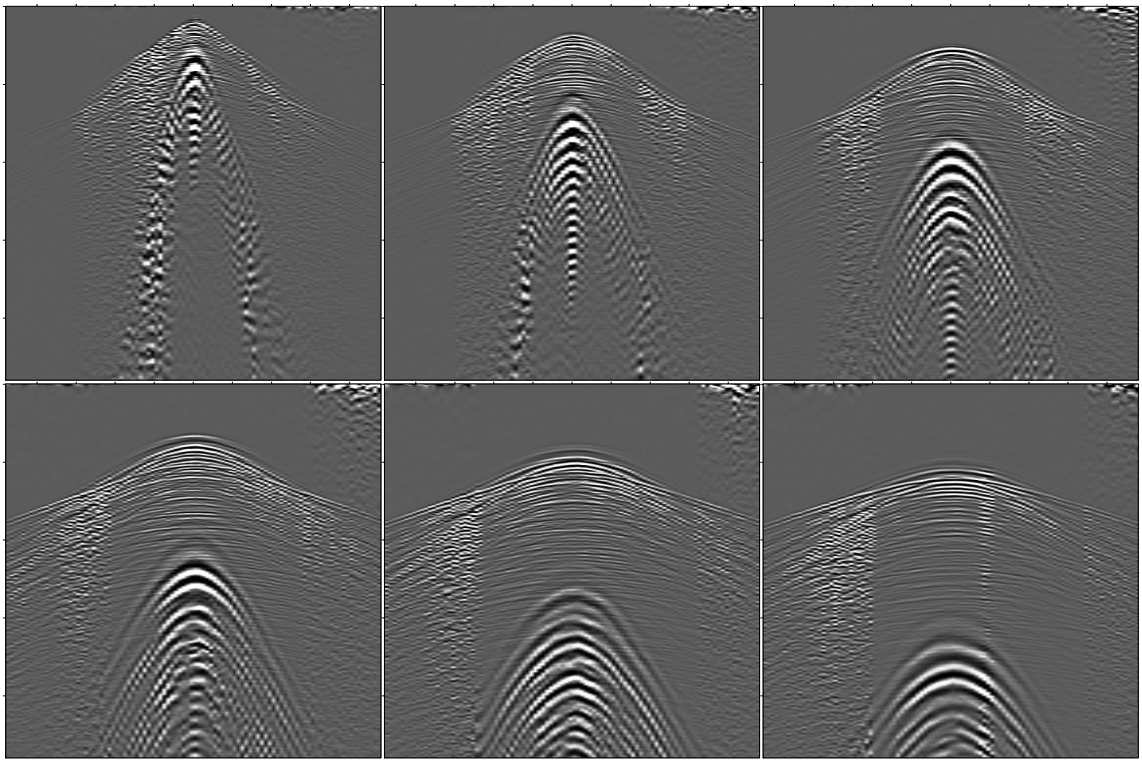

Figure 11: Sand dunes. One shot, six parallel receiver lines.

INVERSION AND NOISE REMOVAL

Here we relate the basic theoretical statement of geophysical inverse theory to the basic theoretical statement of separation of signals from noises.

A common form of linearized **geophysical inverse theory** is

$$\mathbf{0} \approx \mathbf{W}(\mathbf{Lm} - \mathbf{d}) \tag{7}$$

$$\mathbf{0} \approx \epsilon\mathbf{Am} \tag{8}$$

We choose the operator $\mathbf{L} = \mathbf{I}$ to be an identity and we rename the model $\mathbf{m}$ to be signal $\mathbf{s}$. Define noise by the decomposition of data into signal plus noise, so $\mathbf{n} = \mathbf{d} - \mathbf{s}$. Finally, let us rename the weighting (and filtering) operations $\mathbf{W} = \mathbf{N}$ on the noise and $\mathbf{A} = \mathbf{S}$ on the signal. Thus the usual model fitting becomes a fitting for signal-noise separation:

$$0 \approx \mathbf{N}(-\mathbf{n}) = \mathbf{N}(\mathbf{s} - \mathbf{d}) \tag{9}$$

$$0 \approx \epsilon\mathbf{Ss} \tag{10}$$

SIGNAL-NOISE DECOMPOSITION BY DIP

Choose noise $\mathbf{n}$ to be energy that has no spatial correlation and signal $\mathbf{s}$ to be energy with spatial correlation consistent with one, two, or possibly a few plane-wave segments. (Another view of noise is that a huge number of plane waves is required to define the

wavefield; in other words, with **Fourier analysis** you can make anything, signal or noise.) We know that a first-order differential equation can absorb (kill) a single plane wave, a second-order equation can absorb one or two plane waves, etc. In practice, we will choose the order of the wavefield and minimize power to absorb all we can, and call that the signal.

$\mathbf{S}$ is the operator that absorbs (by prediction error) the plane waves and $\mathbf{N}$ absorbs noises and $\epsilon > 0$ is a small scalar to be chosen. The difference between $\mathbf{S}$ and $\mathbf{N}$ is the spatial order of the filters. Because we regard the noise as spatially uncorrelated, $\mathbf{N}$ has coefficients only on the time axis. Coefficients for $\mathbf{S}$ are distributed over time and space. They have one space level, plus another level for each plane-wave segment slope that we deem to be locally present. In the examples here the number of slopes is taken to be two. Where a data field seems to require more than two slopes, it usually means the "patch" could be made smaller.

It would be nice if we could forget about the goal (10) but without it the goal (9), would simply set the signal $\mathbf{s}$ equal to the data $\mathbf{d}$. Choosing the value of ϵ will determine in some way the amount of data energy partitioned into each. The last thing we will do is choose the value of ϵ, and if we do not find a theory for it, we will experiment.

The operators $\mathbf{S}$ and $\mathbf{N}$ can be thought of as "leveling" operators. The method of least-squares sees mainly big things, and spectral zeros in $\mathbf{S}$ and $\mathbf{N}$ tend to cancel spectral lines and plane waves in $\mathbf{s}$ and $\mathbf{n}$. (Here we assume that power levels remain fairly level in time. Were power levels to fluctuate in time, the operators $\mathbf{S}$ and $\mathbf{N}$ should be designed to level them out too.)

None of this is new or exciting in one dimension, but I find it exciting in more dimensions. In seismology, quasi-sinusoidal signals and noises are quite rare, whereas local plane waves are abundant. Just as a short one-dimensional filter can absorb a sinusoid of any frequency, a compact two-dimensional filter can absorb a wavefront of any dip.

To review basic concepts, suppose we are in the one-dimensional frequency domain. Then the solution to the fitting goals (10) and (9) amounts to minimizing a quadratic form by setting to zero its derivative, say

$$0 \quad = \quad \frac{\partial}{\partial \mathbf{s}^{\mathrm{T}}} \left((\mathbf{s}^{\mathrm{T}} - \mathbf{d}^{\mathrm{T}}) \mathbf{N}^{\mathrm{T}} \mathbf{N} (\mathbf{s} - \mathbf{d}) + \epsilon^2 \mathbf{s}^{\mathrm{T}} \mathbf{S}^{\mathrm{T}} \mathbf{S} \mathbf{s} \right) \tag{11}$$

which gives the answer

$$\mathbf{s} \quad = \quad \left(\frac{\mathbf{N}^{\mathrm{T}}\mathbf{N}}{\mathbf{N}^{\mathrm{T}}\mathbf{N} \;+\; \epsilon^2 \mathbf{S}^{\mathrm{T}}\mathbf{S}} \right) \mathbf{d} \tag{12}$$

$$\mathbf{n} \quad = \quad \mathbf{d} - \mathbf{s} \quad = \quad \left(\frac{\epsilon^2 \mathbf{S}^{\mathrm{T}}\mathbf{S}}{\mathbf{N}^{\mathrm{T}}\mathbf{N} \;+\; \epsilon^2 \mathbf{S}^{\mathrm{T}}\mathbf{S}} \right) \mathbf{d} \tag{13}$$

To make this really concrete, consider its meaning in one dimension, where signal is white $\mathbf{S}^{\mathrm{T}}\mathbf{S} = 1$ and noise has the frequency ω_0, which is killable with the multiplier $\mathbf{N}^{\mathrm{T}}\mathbf{N} = (\omega - \omega_0)^2$. Now we recognize that equation (12) is a notch filter and equation (13) is a narrow-band filter.

The analytic solutions in equations (12) and (13) are valid in 2-D **Fourier space** or dip space too. I prefer to compute them in the time and space domain to give me tighter control on window boundaries, but the Fourier solutions give insight and offer a computational speed advantage.

Let us express the fitting goal in the form needed in computation.

$$\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \quad \approx \quad \begin{bmatrix} \mathbf{N} \\ \epsilon \mathbf{S} \end{bmatrix} \mathbf{s} + \begin{bmatrix} -\mathbf{N}\mathbf{d} \\ \mathbf{0} \end{bmatrix} \tag{14}$$

user/gee/signoi.c

```
void signoi_lop (bool adj, bool add, int n1, int n2,
                 float *data, float *sign)
/*< linear operator >*/
{
    sf_helicon_init (nn);
    sf_polydiv_init (nd, ss);

    sf_adjnull(adj,add,n1,n2,data,sign);

    sf_helicon_lop (false, false, n1, n1, data, dd);
    sf_solver_prec(sf_helicon_lop, sf_cgstep, sf_polydiv_lop,
                   nd, nd, nd, sign, dd, niter, eps,
                   "verb", verb, "end");
    sf_cgstep_close();

    nn++;
    ss++;
}
```

As with the missing-data subroutines, the potential number of iterations is large, because the dimensionality of the space of unknowns is much larger than the number of iterations we would find acceptable. Thus, sometimes changing the number of iterations `niter` can create a larger change than changing `epsilon`. Experience shows that helix preconditioning saves the day.

Signal/noise decomposition examples

Figure 12 demonstrates the signal/noise decomposition concept on synthetic data. The signal and noise have similar frequency spectra but different dip spectra.

Before I discovered helix preconditioning, Ray Abma found that different results were obtained when the fitting goal was cast in terms of $\mathbf{n}$ instead of $\mathbf{s}$. Theoretically it should not make any difference. Now I believe that with preconditioning, or even without it, if there are enough iterations, the solution should be independent of whether the fitting goal is cast with either $\mathbf{n}$ or $\mathbf{s}$.

Figure 13 shows the result of experimenting with the choice of ϵ. As expected, increasing ϵ weakens $\mathbf{s}$ and increases $\mathbf{n}$. When ϵ is too small, the noise is small and the signal is almost the original data. When ϵ is too large, the signal is small and coherent events are pushed into the noise. (Figure 13 rescales both signal and noise images for the clearest display.)

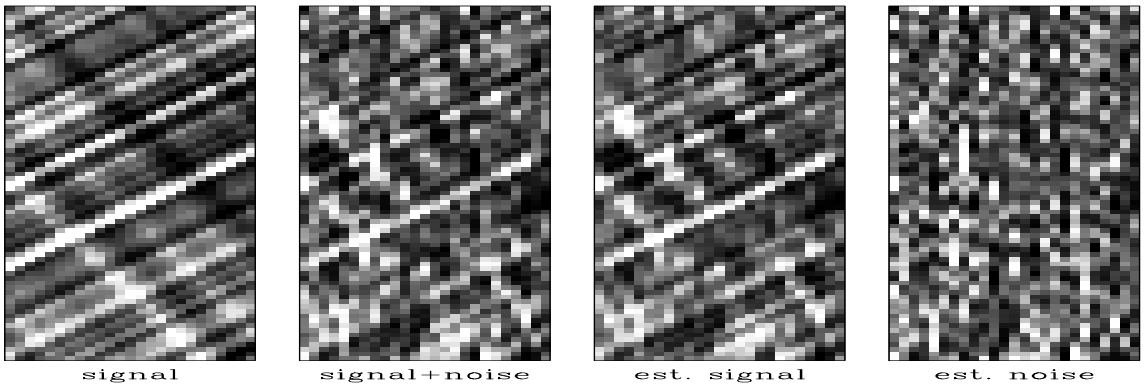

Figure 12: The input signal is on the left. Next is that signal with noise added. Next, for my favorite value of `epsilon=1.`, is the estimated signal and the estimated noise.

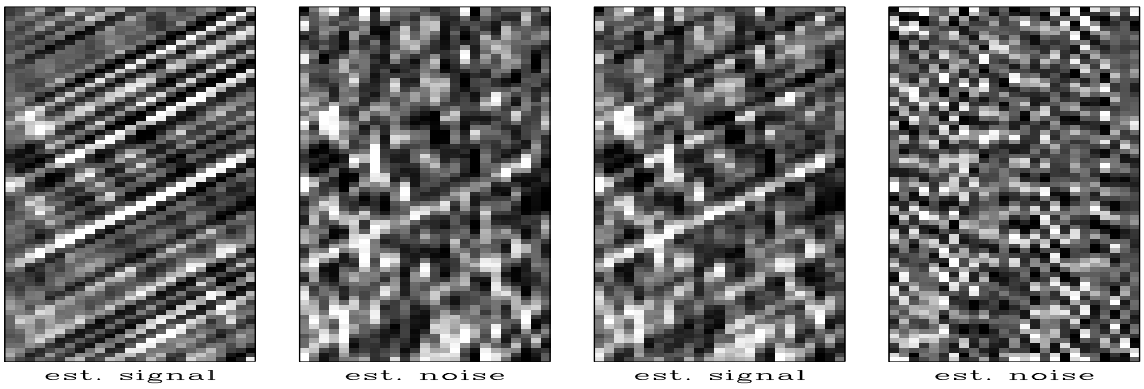

Figure 13: Left is an estimated signal-noise pair where `epsilon=4` has improved the appearance of the estimated signal but some coherent events have been pushed into the noise. Right is a signal-noise pair where `epsilon=.25`, has improved the appearance of the estimated noise but the estimated signal looks no better than original data.

Notice that the leveling operators $\mathbf{S}$ and $\mathbf{N}$ were both estimated from the original signal and noise mixture $\mathbf{d} = \mathbf{s} + \mathbf{n}$ shown in Figure 12. Presumably we could do even better if we were to reestimate $\mathbf{S}$ and $\mathbf{N}$ from the estimates $\mathbf{s}$ and $\mathbf{n}$ in Figure 13.

Spitz for variable covariances

Since signal and noise are uncorrelated, the spectrum of data is the spectrum of the signal plus that of the noise. An equation for this idea is

$$\sigma_d^2 \quad = \quad \sigma_s^2 \; + \; \sigma_n^2 \tag{15}$$

This says resonances in the signal and resonances in the noise will both be found in the data. When we are given σ_d^2 and σ_n^2 it seems a simple matter to subtract to get σ_s^2. Actually it can be very tricky. We are never given σ_d^2 and σ_n^2; we must estimate them. Further, they can be a function of frequency, wave number, or dip, and these can be changing during measurements. We could easily find ourselves with a negative estimate for σ_s^2 which would ruin any attempt to segregate signal from noise. An idea of Simon Spitz can help here.

Let us reexpress equation (15) with prediction-error filters.

$$\frac{1}{\bar{A}_d A_d} \quad = \quad \frac{1}{\bar{A}_s A_s} \; + \; \frac{1}{\bar{A}_n A_n} \quad = \quad \frac{\bar{A}_s A_s \; + \; \bar{A}_n A_n}{(\bar{A}_s A_s)(\bar{A}_n A_n)} \tag{16}$$

Inverting

$$\bar{A}_d A_d \quad = \quad \frac{(\bar{A}_s A_s)\; (\bar{A}_n A_n)}{\bar{A}_s A_s \; + \; \bar{A}_n A_n} \tag{17}$$

The essential feature of a PEF is its zeros. Where a PEF approaches zero, its inverse is large and resonating. When we are concerned with the zeros of a mathematical function we tend to focus on numerators and ignore denominators. The zeros in $\bar{A}_s A_s$ compound with the zeros in $\bar{A}_n A_n$ to make the zeros in $\bar{A}_d A_d$. This motivates the "Spitz Approximation."

$$\bar{A}_d A_d \quad = \quad (\bar{A}_s A_s)\; (\bar{A}_n A_n) \tag{18}$$

It usually happens that we can find a patch of data where no signal is present. That's a good place to estimate the noise PEF A_n. It is usually much harder to find a patch of data where no noise is present. This motivates the Spitz approximation which by saying $A_d = A_s A_n$ tells us that the hard-to-estimate A_s is the ratio $A_s = A_d / A_n$ of two easy-to-estimate PEFs.

It would be computationally convenient if we had A_s expressed not as a ratio. For this, form the signal $\mathbf{u} = \mathbf{A}_n \mathbf{d}$ by applying the noise PEF A_n to the data $\mathbf{d}$. The spectral relation is

$$\sigma_u^2 \quad = \quad \sigma_d^2 / \sigma_n^2 \tag{19}$$

Inverting this expression and using the Spitz approximation we see that a PEF estimate on $\mathbf{u}$ is the required A_s in numerator form because

$$A_u \quad = \quad A_d / A_n \quad = \quad A_s \tag{20}$$

Noise removal on Shearer's data

Professor Peter **Shearer**[1] gathered the earthquakes from the IDA network, an array of about 25 widely distributed gravimeters, donated by Cecil Green, and Shearer selected most of the shallow-depth earthquakes of magnitude greater than about 6 over the 1981-91 time interval, and sorted them by epicentral distance into bins 1° wide and stacked them. He generously shared his edited data with me and I have been restacking it, compensating for amplitude in various ways, and planning time and filtering compensations.

Figure 14 shows a test of noise subtraction by multidip narrow-pass filtering on the **Shearer-IDA stack**. As with prediction there is a general reduction of the noise. Unlike with prediction, weak events are preserved and noise is subtracted from them too.

Besides the difference in theory, the separation filters are much smaller because their size is determined by the concept that "two dips will fit anything locally" (`a2=3`), versus the prediction filters "needing a sizeable window to do statistical averaging." The same aspect ratio `a1/a2` is kept and the page is now divided into 11 vertical patches and 24 horizontal patches (whereas previously the page was divided in 3×4 patches). In both cases the patches overlap about 50%. In both cases I chose to have about ten times as many equations as unknowns on each axis in the estimation. The ten degrees of freedom could be distributed differently along the two axes, but I saw no reason to do so.

The human eye as a dip filter

Although the filter seems to be performing as anticipated, no new events are apparent. I believe the reason that we see no new events is that the competition is too tough. We are competing with the human eye, which through aeons of survival has become is a highly skilled filter. Does this mean that there is no need for filter theory and filter subroutines because the eye can do it equally well? It would seem so. Why then pursue the subject matter of this book?

The answer is 3-D. The human eye is not a perfect filter. It has a limited (though impressive) dynamic range. A nonlinear display (such as wiggle traces) can prevent it from averaging. The eye is particularly good at dip filtering, because the paper can be looked at from a range of grazing angles and averaging window sizes miraculously adjust to the circumstances. The eye can be overwhelmed by too much data. The real problem with the human eye is that the retina is only two-dimensional. The world contains many three-dimensional data volumes. I don't mean the simple kind of 3-D where the contents of the room are nicely mapped onto your 2-D retina. I mean the kind of 3-D found inside a bowl of soup or inside a rock. A rock can be sliced and sliced and sliced again and each slice is a picture. The totality of these slices is a movie. The eye has a limited ability to deal with **movies** by optical persistence, an averaging of all pictures shown in about 1/10 second interval. Further, the eye can follow a moving object and perform the same averaging. I have learned, however, that the eye really cannot follow two objects at two different speeds

[1] I received the data for this stack from Peter Shearer at the **Cecil and Ida Green** Institute of Geophysics and Planetary Physics of the Scripps Oceanographic Institute. I also received his permission to redistribute it to friends and colleagues. Should you have occasion to copy it please reference him. Examples of earlier versions of these stacks are found in the references. Professor Shearer may be willing to supply newer and better stacks. His electronic mail address is `shearer@mahi.ucsd.edu`.

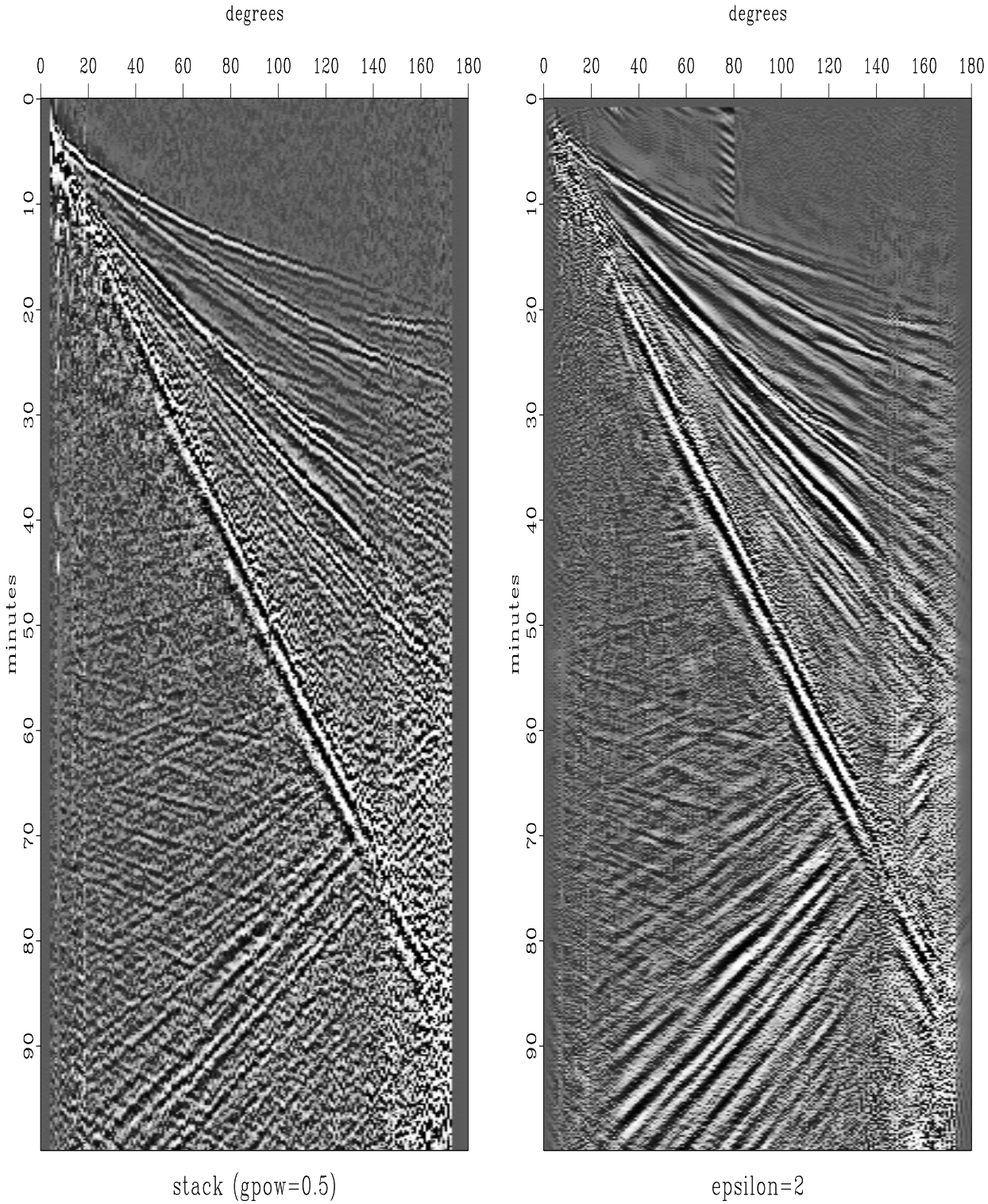

Figure 14: Stack of Shearer's IDA data (left). Multidip filtered (right). It is pleasing that the noise is reduced while weak events are preserved.

and average them both over time. Now think of the third dimension in Figure 14. It is the dimension that I summed over to make the figure. It is the 1° range bin. If we were viewing the many earthquakes in each bin, we would no longer be able to see the out-of-plane information which is the in-plane information in Figure 14.

To view genuinely 3-D information we must see a movie, or we must compress the 3-D to 2-D. There are only a small number of ways to compress 3-D to 2-D. One is to select planes from the volume. One is to sum the volume over one of its axes, and the other is a compromise, a filtering over the axis we wish to abandon before subsampling on it. That filtering is a local smoothing. If the local smoothing has motion (out of plane dip) of various velocities (various dips), then the desired process of smoothing the out of plane direction is what we did in the in-plane direction in Figure 14. But Figure 14 amounts to more than that. It amounts to a kind of simultaneous smoothing in the *two* most coherent directions whereas in 3-D your eye can smooth in only *one* direction when you turn your head along with the motion.

> If the purpose of data processing is to collapse 3-D data volumes to 2-D where they are comprehensible to the human eye, then perhaps data-slope adaptive, low-pass filtering in the out-of-plane direction is the best process we can invent.

My purpose in filtering the earthquake stacks is to form a guiding "pilot trace" to the analysis of the traces *within* the bin. Within each bin, each trace needs small time shifts and perhaps a small temporal filter to best compensate it to . . . to what? to the pilot trace, which in these figures was simply a stack of traces in the bin. Now that we have filtered in the range direction, however, the next stack can be made with a better quality pilot.

SPACE-VARIABLE DECONVOLUTION

Filters sometimes change with time and space. We sometimes observe signals whose spectrum changes with position. A filter that changes with position is called nonstationary. We need an extension of our usual convolution operator `hconest` on page ??. Conceptually, little needs to be changed besides changing `aa(ia)` to `aa(ia,iy)`. But there is a practical problem. Fomel and I have made the decision to clutter up the code somewhat to save a great deal of memory. This should be important to people interested in solving multidimensional problems with big data sets.

Normally, the number of filter coefficients is many fewer than the number of data points, but here we have very many more. Indeed, there are `na` times more. Variable filters require `na` times more memory than the data itself. To make the nonstationary helix code more practical, we now require the filters to be constant in patches. The data type for nonstationary filters (which are constant in patches) is introduced in module `nhelix`, which is a simple modification of module `helix` on page ??. What is new is the integer valued vector `pch(nd)` the size of the one-dimensional (helix) output data space. Every filter output point is to be assigned to a patch. All filters of a given patch number will be the same filter. Nonstationary helixes are created with `createnhelix`, which is a simple modification of module `createhelix` on page ??. Notice that the user must define the `pch(product(nd))` vector before creating a nonstationary helix. For a simple 1-D time-variable filter, presumably

user/gee/nhelix.c

```
typedef struct nhelixfilter {
    int np;
    sf_filter* hlx;
    bool* mis;
    int *pch;
} *nfilter;
```

`pch` would be something like $(1, 1, 2, 2, 3, 3, \cdots)$. For multidimensional patching we need to think a little more.

Finally, we are ready for the convolution operator. The operator `nhconest` on page 25 allows for a different filter in each patch. A filter output `y[iy]` has its filter from the patch `ip=aa->pch[iy]`.

Because of the massive increase in the number of filter coefficients, allowing these many filters takes us from overdetermined to very undetermined. We can estimate all these filter coefficients by the usual deconvolution fitting goal (**??**)

$$\mathbf{0} \quad \approx \quad \mathbf{r} \quad = \quad \mathbf{YKa} + \mathbf{r}_0 \tag{21}$$

but we need to supplement it with some damping goals, say

$$\begin{aligned} \mathbf{0} &\approx \mathbf{YKa} + \mathbf{r}_0 \\ \mathbf{0} &\approx \epsilon\, \mathbf{Ra} \end{aligned} \tag{22}$$

where $\mathbf{R}$ is a roughening operator to be chosen.

Experience with missing data in Chapter **??** shows that when the roughening operator $\mathbf{R}$ is a differential operator, the number of iterations can be large. We can speed the calculation immensely by "preconditioning". Define a new variable $\mathbf{m}$ by $\mathbf{a} = \mathbf{R}^{-1}\mathbf{m}$ and insert it into (22) to get the equivalent preconditioned system of goals.

$$\mathbf{0} \approx \mathbf{YKR}^{-1}\mathbf{m} \tag{23}$$

$$\mathbf{0} \approx \epsilon\, \mathbf{m} \tag{24}$$

The fitting (23) uses the operator $\mathbf{YKR}^{-1}$. For $\mathbf{Y}$ we can use subroutine `nhconest()` on page 25; for the smoothing operator $\mathbf{R}^{-1}$ we can use nonstationary polynomial division with operator `npolydiv()`:

Now we have all the pieces we need. As we previously estimated stationary filters with the module `pef` on page ??, now we can estimate nonstationary PEFs with the module `npef` on page 27. The steps are hardly any different. Near the end of module `npef` is a filter `reshape` from a 1-D array to a 2-D array.

Figure 15 shows a synthetic data example using these programs. As we hope for deconvolution, events are compressed. The compression is fairly good, even though each event has a different spectrum. What is especially pleasing is that satisfactory results are obtained

user/gee/createnhelix.c

```
nfilter createnhelix(int dim      /* number of dimensions */,
                     int *nd      /* data size [dim] */,
                     int *center  /* filter center [dim] */,
                     int *gap     /* filter gap [dim] */,
                     int *na      /* filter size [dim] */,
                     int *pch     /* patching [product(nd)] */)
/*< allocate and output a non-stationary filter >*/
{
    nfilter nsaa;
    sf_filter aa;
    int n123, np, ip, *nh, i;

    aa = createhelix(dim, nd, center, gap, na);

    n123=1;
    for (i=0; i < dim; i++) {
        n123 *= nd[i];
    }
    np = pch[0];
    for (i=0; i < n123; i++) {
        if (pch[i] > np) np=pch[i];
    }
    np++;

    nh = sf_intalloc(np);
    for (ip=0; ip < np; ip++) {
        nh[ip] = aa->nh;
    }
    nsaa = nallocate(np, n123, nh, pch);
    for (ip=0; ip < np; ip++) {
        for (i=0; i < aa->nh; i++) {
            nsaa->hlx[ip]->lag[i] = aa->lag[i];
        }
        nbound(ip, dim, nd, na, nsaa);
    }

    sf_deallocatehelix(aa);

    return nsaa;
}
```

user/gee/nhconest.c

```
    for (iy=0; iy < ny; iy++) {
        if (aa->mis[iy]) continue;

        ip = aa->pch[iy];
        lag = aa->hlx[ip]->lag;
        na = aa->hlx[ip]->nh;

        for (ia=0; ia < na; ia++) {
            ix = iy - lag[ia];
            if (ix < 0) continue;

            if (adj) {
                a[ia+nhmax*ip] += y[iy] * x[ix];
            } else {
                y[iy] += a[ia+nhmax*ip] * x[ix];
            }
        }
    }
```

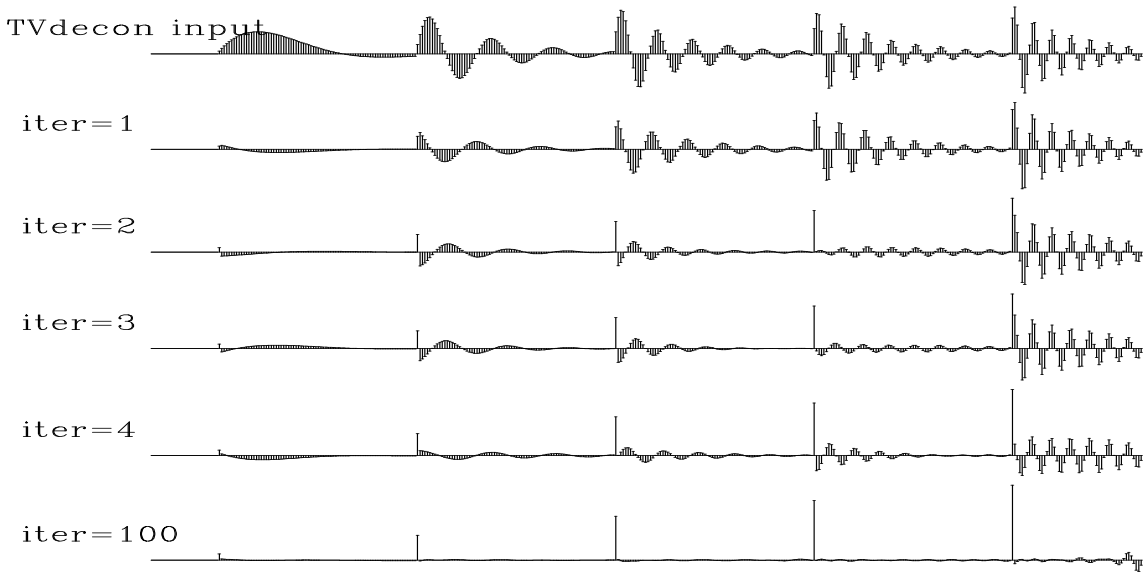

Figure 15: Time variable deconvolution with two free filter coefficients and a gap of 6.

user/gee/npolydiv.c

```
    for (id=0; id < nd; id++) {
        tt[id] = adj? yy[id]: xx[id];
    }

    if (adj) {
        for (iy=nd-1; iy >= 0; iy--) {
            ip = aa->pch[iy];
            lag = aa->hlx[ip]->lag;
            flt = aa->hlx[ip]->flt;
            na = aa->hlx[ip]->nh;
            for (ia=0; ia < na; ia++) {
                ix = iy - lag[ia];
                if (ix < 0) continue;
                tt[ix] -= flt[ia] * tt[iy];
            }
        }
        for (id=0; id < nd; id++) {
            xx[id] += tt[id];
        }
    } else {
        for (iy=0; iy < nd; iy++) {
            ip = aa->pch[iy];
            lag = aa->hlx[ip]->lag;
            flt = aa->hlx[ip]->flt;
            na = aa->hlx[ip]->nh;
            for (ia=0; ia < na; ia++) {
                ix = iy - lag[ia];
                if (ix < 0) continue;
                tt[iy] -= flt[ia] * tt[ix];
            }
        }
        for (id=0; id < nd; id++) {
            yy[id] += tt[id];
        }
    }
```

user/gee/npef.c

```
void find_pef(int nd      /* data size */,
              float *dd   /* data */,
              nfilter aa  /* estimated filter */,
              nfilter rr  /* regularization filter */,
              int niter   /* number of iterations */,
              float eps   /* regularization parameter */,
              int nh      /* filter size */)
/*< estimate non-stationary PEF >*/
{
    int ip, ih, na, np, nr;
    float *flt;

    np = aa->np;
    nr = np*nh;
    flt = sf_floatalloc(nr);

    nhconest_init(dd, aa, nh);
    npolydiv2_init( nr, rr);

    sf_solver_prec(nhconest_lop, sf_cgstep, npolydiv2_lop,
                   nr, nr, nd, flt, dd, niter, eps, "end");
    sf_cgstep_close();
    npolydiv2_close();

    for (ip=0; ip < np; ip++) {
        na = aa->hlx[ip]->nh;
        for (ih=0; ih < na; ih++) {
            aa->hlx[ip]->flt[ih] = -flt[ip*nh + ih];
        }
    }

    free(flt);
}
```

in truly small numbers of iterations (about three). The example is for two free filter coefficients $(1, a_1, a_2)$ per output point. The roughening operator $\mathbf{R}$ was taken to be $(1, -2, 1)$ which was factored into causal and anticausal finite difference.

I hope also to find a test case with field data, but experience in seismology is that spectral changes are slow, which implies unexciting results. Many interesting examples should exist in two- and three-dimensional filtering, however, because reflector dip is always changing and that changes the *spatial* spectrum.

In multidimensional space, the smoothing filter $\mathbf{R}^{-1}$ can be chosen with interesting directional properties. Sergey, Bob, Sean and I have joked about this code being the "double helix" program because there are two multidimensional helixes in it, one the smoothing filter, the other the deconvolution filter. Unlike the biological helixes, however, these two helixes do not seem to form a symmetrical pair.

EXERCISES:

1 Is `nhconest` on page 25 the inverse operator to `npolydiv` on page 26? Do they commute?

2 Sketch the matrix corresponding to operator `nhconest` on page 25. HINTS: Do not try to write all the matrix elements. Instead draw short lines to indicate rows or columns. As a "warm up" consider a simpler case where one filter is used on the first half of the data and another filter for the other half. Then upgrade that solution from two to about ten filters.