

GIPSY MEMO #1

GDSINP

(now and in the future)

by

Kor G. Begeman and Wim Zwitser

This document tries to get rid of some of the unclarities which still seem to exist concerning the input to the SET= keyword of the new GIPSY applications. Also a suggestion for improvement of the system is provided by the authors.

Chapter 1

The set and subset terminology

A set is an N-dimensional data structure which has N different axes names. A subset is an M-dimensional structure within this set, where M can be in the range from 0 to N. So a subset of dimension 0 is one pixel, a subset of dimension N is the whole data set. When the dimension of the subset is less than the dimension of the set, there are more subsets in a set. In fact, the total number of subsets can be expressed as:

$$\text{NSUB} = \prod_{k=M+1}^N \text{NAXIS}_k$$

where NAXIS_k is the length of a dimension not in the subset.

Some application programs allow the user to enter more than one subset, which means that the operation they perform is repeated for each subset. An example of such an application is LOOK, which displays a 2-dimensional subset on a display device.

For the sake of simplicity let us assume throughout this chapter that we have a 3-dimensional dataset with name IN3, which has axes:

AXIS#1 RA-NCP from -63 to 64

AXIS#2 DEC-NCP from -63 to 64

AXIS#3 FREQ-OHEL from 1 to 63

So in order to display a sequence of frequency channels, the user may type

LOOK SET=IN3 FREQ 21:40

This means that he defined his subsets to be 2-dimensional (which is required by LOOK), and that the axis outside(!) the subset is FREQ. In fact, you could call this the repeat axis for which the same operation is repeated. The first axis of the subset is RA, and the second axis of the subset is DEC. If the user wants to look at some LV maps (old-fashioned terminology for RA-FREQ maps), he may specify

LOOK SET=IN3 DEC -15:16

Now the repeat axis is DEC, and the first subset axis is RA and the second is FREQ. This handy feature of GDS allows the user to look at any plane in his 3-dimensional set without having to transpose his data set. However, somewhere you must pay for this and that is in the time it takes to display a map. It will take much more time to display a DEC-FREQ map of IN3 than a RA-DEC map.

The reason for this is simply explained. The data are stored sequentially on disk. If you go from the first datum on disk to the second, the grid position along the first axis is incremented by one. If you go on further and further, you will find that the grid position along the first axis changes fastest (each step), the grid position along the second axis changes less fast (each NAXIS1 step) and along the last axis the change is the slowest (each $\prod_{k=1}^{N-1} \text{NAXIS}_k$ step). So the data along the first axis is retrieved faster than data along any other axis. This is why LOOK cannot display a DEC-RA map or a FREQ-RA map because GDSINP will make the first subset axis the fastest axis and the second subset axis the second fastest axes etc.

Most programs allow the user to select a part of the subset to do the operation on. This is done with the LINE=, AREA=, CUBE=, etc. keywords. These keywords expect two positions which denote a continuous line, area, cube etc. within the subset.

Now suppose that a student named S. Body has to give a couple of demos, and Prof. S. tells him to show some RA-FREQ maps. Now Mr. Body is a smart guy, so beforehand he transposes his data cube with the program COPY. He does this by typing

COPY SET=IN3 DEC -63:64 AREA=-63,1,64,63 SETOUT=OUT3

and ends up with an RA-FREQ-DEC cube. COPY interpretes the input to the SET= keyword just as LOOK does: it copies 128 RA-FREQ subsets to OUT3, the first subset axis becomes the first axis of OUT3, the second subset axis becomes the second axis of OUT3, ..., then the first repeat axis becomes the $(M + 1)^{\text{th}}$ axis of OUT3 and the last repeat axis becomes the last axis of OUT3. So if Prof. S. wants his student to show FREQ-RA maps, this poor student has to run

COPY SET=IN3 RA -63:64 DEC -63:64 LINE=1,63 SETOUT=OUT3

Note that now the dimension of the subset is one less than in the first example. The result of this operation is that student S. Body can now at maximum speed display his FREQ-RA maps, and Prof. S. presents him a bottle of the finest Barolo for his insight into this matter.

In the next chapter we will visit a parallel world, called ITALY, where everybody has nothing else to do than wait and being proud of the most purest GIPSY system they developed.

Chapter 2

How it should be

In ITALY, one of the parallel worlds where you only find purists, a nincompoop called N. Body had to write a program called MOMENTS and a program called MEAN for Prof. S. Body, who got kicked out of the mathematics department before he decided to study astronomy, defined the problem in a pure mathematical way. The program MOMENTS should calculate the zeroed and first moment along a line. Then the moments are defined as follows:

$$M0 = \int_1^{NAXIS1} data(n)dn$$
$$M1 = \frac{\int_1^{NAXIS1} data(n)ndn}{M0}$$

For the program MEAN he decided that it should be able to cope with any number of axes, so he defined:

$$MEAN = \frac{\int_1^{NAXIS1} \dots \int_1^{NAXISM} data(n_1, \dots, n_m)dn_1 \dots dn_m}{NAXIS1 \dots NAXISM}$$

When these program were ready, Prof. S. started to work rightaway and wanted to use MOMENTS to calculate a velocity field from his data in set IN3. He knew how LOOK and COPY worked, and since everything is pure in ITALY, he knew immediately how to run MOMENTS. He sat behind his favorite workstation and typed

MOMENTS SET=IN3 RA -63:64 DEC -63:64 LINE=1,63 SETOUT=OUT3

After a while he ran disk and found that MOMENTS had created a 3-dimensional set (OUT3) with

```
AXIS#1 RA-NCP    from -63 to 64
AXIS#2 DEC-NCP   from -63 to 64
AXIS#3 PARAM     from  0 to  1
```

Encouraged by this he started MEAN to create a continuum map, the average of channels 2 to 15 for a similar set like IN3 named IN3-2. For this he typed

MEAN SET=IN3-2 RA -63:64 DEC -63:64 LINE=2,15 SETOUT=OUT3-2

After a while he ran DISK again to find that MEAN had created a 2-dimensional set with

AXIS#1 RA-NCP from -63 to 64

AXIS#2 DEC-NCP from -63 to 64

And when Prof. S. is still alive he will still be waiting for his velocity field and continuum map.

In the next chapter we go back to our own parallel world, which is called PARALLEL in case you didn't know it, and see how parallel student S. Body solves this problem.

Chapter 3

The parallel world

In the parallel world PARALLEL student S. Body had to write the same programs MOMENTS and MEAN for his Prof. S. Since in PARALLEL time is money and you had to write at least one paper a week, student Body got the message that he had to write these programs as efficiently as possible. After a little while Body came up with the following idea: If Prof. S. wants to use MOMENTS to calculate a velocity field and MEAN to create a continuum map, then the most efficient way to do this is in parallel mode. Instead of calculating the moments for one line at a time, it would be much faster to do it for more lines simultaneously. The gain in speed would arise from the way in which the data are stored. Usually the parallel direction would be much faster than the direction of the operation (for MOMENTS the integration direction).

Body realized that this means that program MOMENTS is not pure anymore. The axes names Prof. S. would have to type in after the set name would not be called repeat axes, but operation axes (student Body could not think of a better word for it). Likewise for MEAN, which Body would allow to average in more than one direction, the axes of operation would have to be typed in after the set name. So, this is exactly the reverse of how you would do it in ITALY.

When these programs were ready, Prof. S. immediately sat behind his workstation and typed

```
MOMENTS SET=IN3 FREQ 1:63 AREA=-63,-63,64,64 SETOUT=OUT3
```

Two minutes later he heard the program ready beep and found that MOMENTS had created a set with axes

```
AXIS#1 RA-NCP from -63 to 64
```

```
AXIS#2 DEC-NCP from -63 to 64
```

```
AXIS#3 PARAM from 0 to 1
```

Next he tried MEAN to create a continuum map by typing

```
MEAN SET=IN3-2 FREQ 2:15 45:59 AREA=-63,-63,64,64  
SETOUT=OUT-3
```

It took MEAN only one minute to create the continuum map. Prof. S. made a nice contour plot of the velocity field and continuum map, and two days later he had his paper submitted.

In the mean time, student Body was still pondering about the problem of different types of applications. He found that most programs could be split up in two groups. One group he called class 1 programs, which operate like LOOK (CLEAN, CPLOT, COMBIN, STAT) and have repeat axes, the other group he called class 2 and operate like MOMENTS (GAUFIT, GAUSS, MEAN, SUM) and have operation axes. The programs in class 2 could all be speeded up if they would use the same trick as he applied to MOMENTS. But then the question arises whether it would be possible to have for class 2 programs mixed input of operation axes and repeat axes. In other words, should an operation as performed by MOMENTS be repeatable.

In the next and last chapter we will present our new proposal.

Chapter 4

Conclusion

For the sake of clarity the ITALY solution is best without doubt, but in performance the PARALLEL solution clearly wins. This is true for radio data, but is it true for other applications? Suppose that an optical astronomer comes with some slit spectra, LAMBDA being the first axis and position being the other axis. He would not gain at all from the PARALLEL solution of MOMENTS. He would be able to use MOMENTS of course in both worlds, but to him the PARALLEL solution might seem weird.

And what about the mix of repeat axes and operation axes. In general one would gain little if class 2 programs were also capable of repeating operations, and they loose a lot in clarity and simplicity to the user. At the moment there are 9 GIPSY applications which are in class 2. They are AVRAGE, GAUFIT, GAUSS, MEAN, PROFIL, PROFIT, SUM, VEL, VELSMO.

Therefore we propose to use either repeat axes or operation axes, not a mixture of both. This means that for program MOMENTS you may only enter one axis name (being the operation axes), and for program MEAN you would be allowed to enter as many operation axes as you like. Notice that for class 2 programs there might be a restriction on the number of axes outside the subset, while for class 1 programs there might be a restriction for the number of axes inside the subset. It is clear that the documentation should state very clearly whether a program is a class 1 or a class 2 program, to avoid confusion (if possible).

Just to give an example of how it then would look like, typing

```
MEAN SET=IN3 RA -63:64 DEC -63:64 LINE 1,63 SETOUT=OUT3
```

would mean that OUT3 contains an average spectrum (averaged over all sky positions), and typing

```
MOMENTS SET=IN3 FREQ 1:63 AREA=-63,-63,64,64 SETOUT=OUT3
```

would give you the wanted velocity field. Note that if IN3 would have an extra axis (let's say POLN), then MOMENTS would ask for CUBE= and not AREA=. This is a consequence of not mixing repeat and operation axes.