An overview of `rlpSpec`: Adaptive sine multitaper power spectral density estimation

Andrew J. Barbour <andy.barbour@gmail.com> and Robert L. Parker

January 23, 2013

Abstract

The purpose of this vignette is to provide an overview of the features included in `rlpSpec`, which allow the user to compute sophisticated power spectral density (PSD) estimates for a univariate series, with very little tuning effort. The sine multitapers are used in which the number of tapers varies with spectral shape, according to the optimal value proposed by Riedel and Sidorenko [1995]. The adaptive procedure iteratively refines the optimal number of tapers at each frequency, which, assuming convergence, will produce spectra with significantly reduced variance (compared to naïve estimators), and minimum biasing effects. Resolution and uncertainty in a multitaper scheme are controlled by the number of tapers used. This means we do not need to resort to either windowing methods which inherently degrade resolution of low-frequency features (e.g. Welch), or smoothing kernels which can badly distort important features without careful tuning (e.g. Daniell, as in `stats::spec.pgram`). In this sense `rlpSpec` is best suited for data having spectra with both large dynamic range and strong, sharply changing features.

Contents

1 Quick start: A minimal example. **1**

2 Comparisons with other methods **4**
- 2.1 `stats::spectrum` 4
- 2.2 `multitaper::` 5

3 Assesing spectral properties **5**

4 Call overview **6**

1 Quick start: A minimal example.

First load the package into the namespace:

```
> library(rlpSpec)
```

We now need a dataset to analyze. Among the datasets included in `rlpSpec` is a subset of the Magnetic Satellite (MAGSAT) mission [Langel et al., 1982]. Specifically, we have included along-track measurements of horizontal magnetic-field strength from a gimballed, airborne magnetometer, sampled once every kilometer, which means the spectrum may represent crustal magnetization with wavelengths longer than 2 km.

```
> data(magsat)
```

The format of the data set is a `data.frame` with four sets of information:

```
> names(magsat)

[1] "km"    "raw"   "clean" "mdiff"
```

The `raw` and `clean` names represent raw and edited intensities respectively, expressed in units of nanoTesla; `mdiff` is the difference between them. The difference between them is a matter of just a few points attributable to instrumental malfunction.

```
> subset(magsat, abs(mdiff)>0)

    km    raw   clean     mdiff
403  0  209.1 -3.6355 -212.7355
717  0 -248.7 -9.7775  238.9225
```

These deviations can, as we will see, adversely affect the accuracy of any PSD estimate, multitaper or otherwise.

Setting aside any discussion regarding sample stationarity, we can find power spectral density (PSD) estimates for the two series quite simply:

```
> psdr <- pspectrum(magsat$raw)
> psdc <- pspectrum(magsat$clean)
```

Each `pspectrum` command calculates a pilot PSD, followed by four iterations of refinement (the default). With each iteration the number of tapers is adjusted to the optimal number, based on the weighted spectral derivatives, following Riedel and Sidorenko [1995]. In general, spectral variance is reduced with sequential refinements, but is not necessarily quaranteed to converge. Note that in the example the sampling frequency of both series is km^{-1}, so we need not change the sampling rate argument.

Let's now visualize the two PSD estimates, recalling that the difference between the `raw` and `clean` samples is a mere two points. [1] Figure 1 compares the spectra for the `raw` and `clean` samples. This plot shows a drastic improvement in shape between the two series, simply because the large outliers have been

[1] Note that `pspectrum` returns an object with class `spec`, so we have access to methods within `stats`, including `plot.spec`.

```
> plot(psdc, log="dB")
> lines(psdr$freq, 10*log10(psdr$spec), col="red")
> legend("bottomleft",c("magsat$raw","magsat$clean"),col=c("red","black"),lty=1,lwd=2)
```

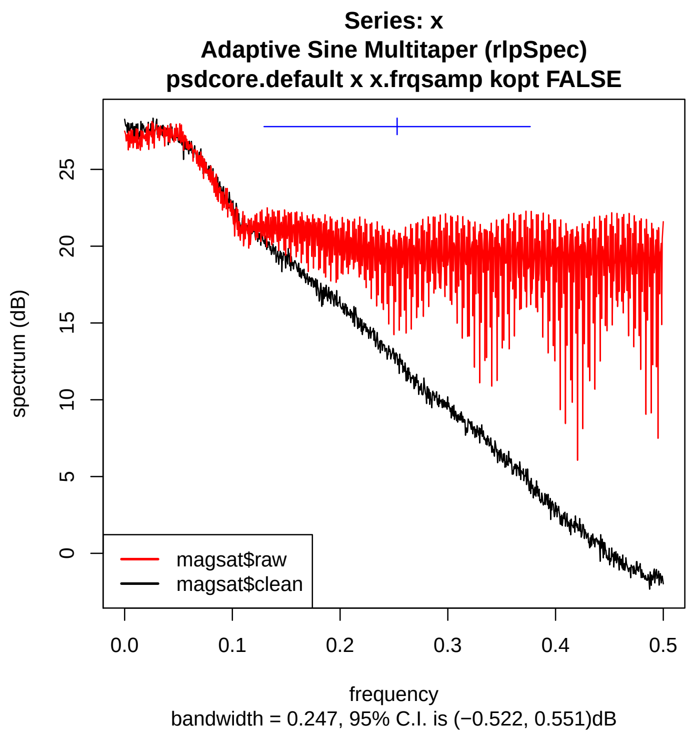

Figure 1: Comparison of power spectral densities for the `MAGSAT` data included with `rlpSpec`.

removed. The `clean` PSD shows the very red spectrum typical of geophysical processes [Agnew, 1992]. It also shows a rolloff in signal somewhere around the 20 kilometer wavelengths; whereas, the `raw` PSD looks highly unrealistic at higher wavelengths, and shows some curvature bias at low frequencies.

2 Comparisons with other methods

As we have shown in the MAGSAT example, improved understanding of the physics behind the signals in the data is of great concern. Assuming a sample is free of non-physical points, how do PSD estimates from `rlpSpec` compare with other methods?

2.1 `stats::spectrum`

Included in the core distribution of R is `stats::spectrum`, which accesses `stats::spec.ar` or `stats::spec.pgram` for either parametric and non-parametric estimation, respectively. Our method is non-parametric; hence, we will compare to the latter.

Included in `rlpSpec` is an option to compare the results with a naïve estimator–the raw periodogram–from within the spectrum calculator, `psdcore`. In R this estimator is equivalent to running:

```
> spec.pgram(X, pad=1, taper=0, detrend=FALSE, demean=FALSE, plot=F)
```

which `psdcore` also calculates.

As a matter of bookkeeping, we must deal with the working environment accessed by `rlpSpec` functions. Specifically, `psdcore` does not clear anything from the working environment, but accesses some of its content under certain conditions; hence, we must first manually clear the environment to prevent bogus results:

```
> str(rlp_envStatus())

List of 5
 $ env_name        : chr ".rlpSpecEnv"
 $ obviously_exists: logi TRUE
 $ listing         : chr [1:11] "fft_even_demeaned_padded" "histlist" "init" "len_even" ...
 $ env_init        : chr "refreshed at 2013-01-23 23:04:01"
 $ env_status_stamp: POSIXct[1:1], format: "2013-01-23 23:04:02"

> rlpSpec:::rlp_envClear()
> str(rlp_envStatus())

List of 5
 $ env_name        : chr ".rlpSpecEnv"
 $ obviously_exists: logi TRUE
 $ listing         : chr "init"
```

```
 $ env_init        : chr "refreshed at 2013-01-23 23:04:02"
 $ env_status_stamp: POSIXct[1:1], format: "2013-01-23 23:04:02"
```

and then re-calculate the multitaper PSD and the raw periodogram. The results are shown in Figure 2.1.

```
> ntap <- psdc$taper
> psdcore(magsat$clean, ntaper=ntap, plotpsd=TRUE)
```

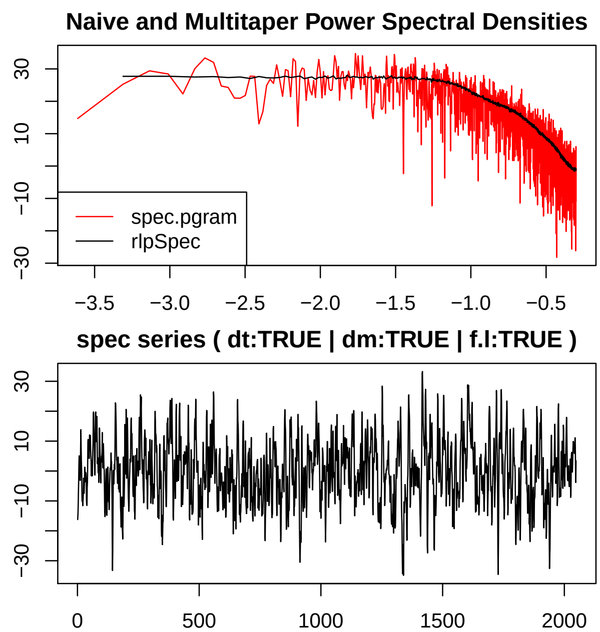

Figure 2: Top: Comparison between naïve and multitaper PSD estimators for the clean `MAGSAT` data. The frequency axis is in units of $\log_{10}$ km^{-1}, and power axis is in decibels. Bottom: The spatial series used to estimate the PSDs.

2.2 `multitaper::`

3 Assesing spectral properties

In a multitaper scheme, the computation of resolution and uncertainty (shown as blue lines in Figure 1 depends on the the number of tapers; hence, the methods internal to `plot.spec` are not appropriate.

4 Call overview

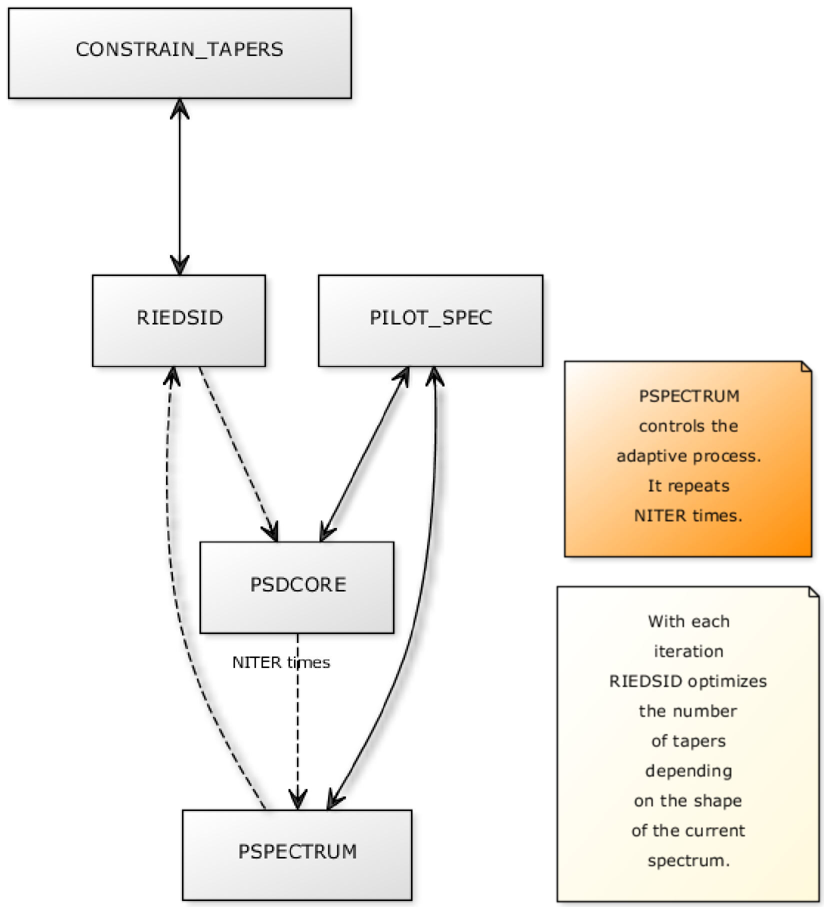

Figure 3: Simplified call graph.

References

D.C. Agnew. The time-domain behavior of power-law noises. *Geophysical Research Letters*, 19:333–336, 1992.

R. Langel, G. Ousley, J. Berbert, J. Murphy, and M. Settle. The MAGSAT mission. *Geophysical Research Letters*, 9(4):243–245, 1982.

K.S. Riedel and A. Sidorenko. Minimum bias multiple taper spectral estimation. *Signal Processing, IEEE Transactions on*, 43(1):188–195, 1995.