Instruction for `GrplassoSeq`

Using the R command-line, the package can be installed by the following command at the directory where the downloaded file exists:

```
> install.packages(c("grplasso","seqinr","foreach")) # if necessary
> install.packages("GrplassoSeq_1.0.tar.gz", repos = NULL, type = "source")
```

A set of sequences and their wavelengths should be provided as two separated input files. A sequence data should be the FASTA format which can be read by `read.fasta` function of the R `seqinr` package[1]. Note that the sequences should be aligned into the same length beforehand including both of training and target proteins. Wavelengths should be stored in a text file in which each line contains a wavelength for a sequence. The order of the sequences in the FASTA file and the wavelength file should be consistent. For wavelength unknown sequences, please specify "NA" in the wavelength file. Proteins specified as "NA" are regarded as target proteins, and other proteins with wavelengths are used as training proteins.

The following command performs the group LASSO optimization using the training proteins:

```
> result <- train_glasso("your_FASTA_file","your_wavelength_file")
```

Then, the prediction for the target proteins can be obtained by

```
> result$predicted_wavelen
```

The regression coefficients can be accessed through the object created by the `grplasso`[2] package:

```
> coef(result$model)
```

Instead of using all residues, a subset of positions in the aligned sequence can be specified to focus on some region of interest in the sequences. The positions should be written in another text file in which each line contains a position index in $\{1, \ldots, N\}$. The third (optional) argument of `train_glasso` function is for the position file:

```
> result <- train_glasso("your_FASTA_file",
                         "your_wavelength_file","your_position_file")
```

For more detail, see the instruction in the package.

[1] `https://cran.r-project.org/web/packages/seqinr/index.html`
[2] https://cran.r-project.org/package=grplasso