Compression Algorithm for Infrared Hyperspectral Sounder Data†

Irina Gladkova[1], Leonid Roytman[1], Mitch Goldberg[2]

[1] NOAA-CREST, CCNY [2] NOAA-NESDIS

Abstract

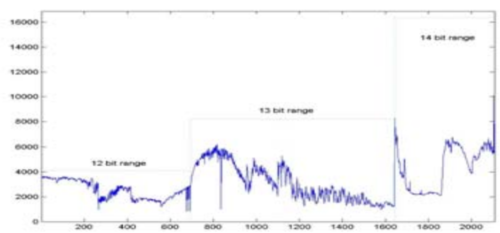
In this poster, we present a new lossless algorithm for compression of the signals from NOAA's environmental satellites. We are using current spacecraft to simulate data from NOAA's environmental satellites and focusing on Aqua Spacecraft's AIRS isntrument in our case study.

Our algorithm uses the PCA method in conjunction with an adaptive clustering procedure developed specifically for the considered sounder data. The goal of the clustering procedure is tp transform the data set into clusters, each with distributions that are close to normal. After these clusters are identified, the standard PCA can then be applied. The negligible memory increase (of about 0.03%) due to the clustering procedure justifies the use of PCA, which is known to be an optimal linear projection algorithm for the normally distributed data.

Algorithm at a Glance

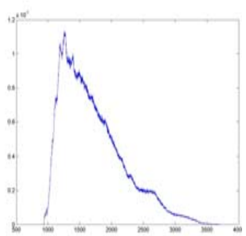
a) Channel Partitioning

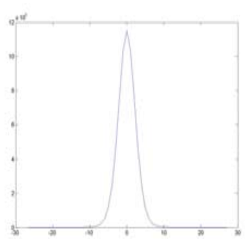
During the first stage, we partition the data into three units. The partitioning takes into account that the range of the digital counts varies (12, 13, and 14 bits) with respect to the channel index (frequency), and hence we process each of these ranges separately.



b) Adaptive Clustering

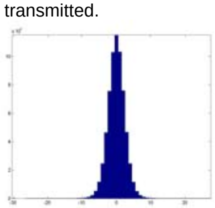
The purpose of this step is to transform the data so that its distribution is as close to normal as possible.

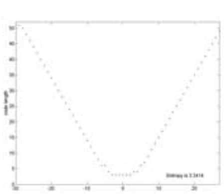




c) Approximated Entropy Coding

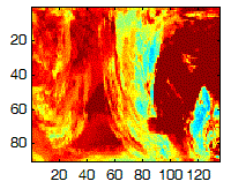
We build our Huffman codebook based on a normal distribution with variance computed from the residuals so that the codebook need not be transmitted.

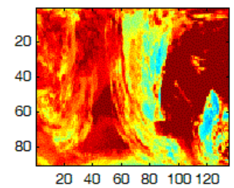




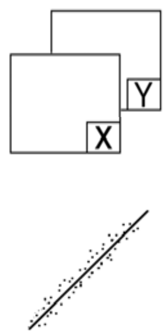
Adaptive Clustering

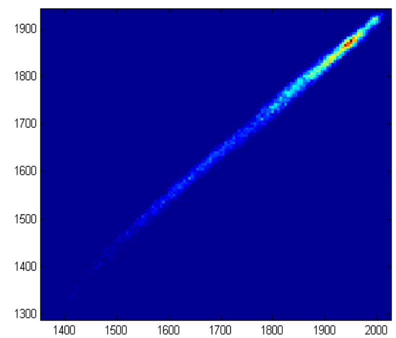
To justify the need for clustering, we start with an observation about the nature of the correlation of data points between consecutive satellite images. As we progress from one image to the next, as illustrated in the images below, we see that they, like many other images in the granule, are visually similar.

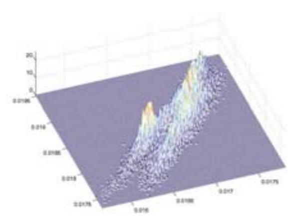


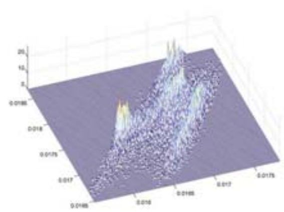


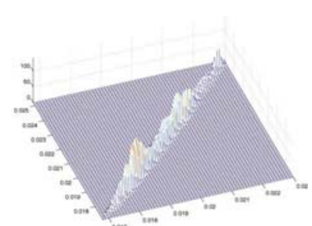
To illustrate this relationship, one can plot the gray level values of pairs of consecutive images. As can be noted from the scatter plots below, the points of the actual data have more than one concentration (cluster), and different clusters have different mean and principal projection directions.

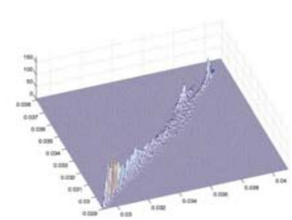












Mathematical Formulation

We formulate this clustering problem in terms of a non-linear least squares optimization problem, which identifies clusters based on the distances to the principal direction.

$$L(A_0,A_1,\dots,A_m)=\sum_{r\in D}\min_{0\le b\le m}\varepsilon(A_b,r),$$

The objective function expresses our criterion for a 'point', i.e. a matrix, to correspond to a linear subspace of the type required,i.e., its columns are a basis for such a subspace. But this is not a smooth function, so we will modify it in the following way:

$$L(A_0,A_1,\dots,A_m)=-\lambda^{-1}\sum_{r\in D}\ln\sum_{b=0}^{m}\phi_{r,b},\quad \lambda\to+\infty, \tag{2}$$

where

$$\phi_{r,b}=\exp(-\lambda\varepsilon(A_b,r)). \tag{3}$$

But

$$\sum_{b=0}^{m}\phi_{r,b}\approx\phi_{r,\beta},\quad \text{where}\quad \varepsilon(A_\beta,r)=\min_{0\le b\le m}\varepsilon(A_b,r).$$

$\varepsilon(A_b,r)$ can be represented as the ratio of two determinants:

$$\varepsilon(A_b,r)=\frac{\det(\widetilde{A_b^*},\widetilde{A_b})}{\det(A_b^*,A_b)},$$

So the considered Lagrangian (2) is a smooth function and we will differentiate it with respect to the *k* th column *Ab;k* of the matrix *Ab* in order to obtain the first order condition for optimization:

$$\frac{\partial L}{\partial A_{b,k}}\equiv\sum_{r\in D}\frac{\partial\varepsilon(A_b,r)}{\partial A_{b,k}}\varphi_{r,b}=0,$$

where

$$\varphi_{r,b}=\frac{\phi_{r,b}}{\phi_{r,0}+\ldots+\phi_{r,m}},$$

Results

Granule	Location	Ratio
9	Pacific Ocean, Daytime	3.3314
16	Europe, Nighttime	3.3291
60	Asia, Daytime	3.2315
82	North America, Nighttime	3.3651
120	Antarctica, Nighttime	3.2831
126	Africa, Daytime	3.2259
129	Arctic, Daytime	3.3711
151	Australia, Nighttime	3.2162
182	Asia, Nighttime	3.1771
193	North America, Daytime	3.2278

* Data consists of AIRS granules in the form of digital counts.

† *Presented at Data Compression Conference, Snowbird, Utah, Mar 28-31 2005, Sponsored by NOAA/NESDIS under Roger Heymann (OSD), Tim Schmit (ORA), HES Compression Group*