Sparse Matrix Techniques for Coupling Independent Hydrological and Meteorological Models

Carlie J. Coats, Jr.
Baron Advanced Meteorological Systems, LLC,
Research Triangle Park, NC 27709-5064

Abstract

Grid-to-grid, grid-to-watershed, and similar geospatial transforms can very efficiently be realized by sparse matrix technology. Such transforms are particularly useful for the grid cell to/from watershed mappings needed for coupling watershed-based hydrological models with grid-cell based meteorological models. We have used these transforms for coupling the TOPLATS hydrological model [Peters-Lidard et al 1997, 1999] with both the MM5 meteorological model and the NEXRAD Stage IV precipitation analyses, and for various other model couplings. We describe the advantages of this technique, as compared with other techniques, which frequently are purely grid-based.

Concept

In the coupled modeling of disparate media, the natural spatial and temporal scales of the medial being modeled frequently are quite different. This occurs, for example, when coupling hydrological and meteorological models: the spatial structure of the hydrological models is defined by the catchment structure or by a refinement of that structure used to model sub-catchment behavior. The spatial scale of the latter is frequently measured in tens of meters, whereas the temporal scale ranges from tens of minutes up to days. The spatial scale of a meteorological model is defined by its grid structure, typically a rectangular grid with a resolution of a few kilometers, with an accompanying temporal scale measured in seconds or minutes. The coupling of hydrological and meteorological models requires that variables on either of these geospatial structures to be re-gridded onto the other structure. The very high resolutions required (whether the temporal resolution, in the case of meteorology, or the spatial resolution, for hydrology) require computational efficiency of the algorithms used. Fortunately, for the most part, the grids themselves are time-invariant; this allows us to eliminate what would be much redundant computation.

For couplings to be conservative of the relevant physical laws, the grid-to-grid transformations generally must be linear. Moreover, "effects are local", i.e., the value for a particular cell in the output grid only depends upon the values of those input-grid cells that intersect it. This means that these grid-to-grid transformations fit the conditions of the *Representation Theorem* from abstract linear algebra: one may specify a “standard order” for the cells of the input and output grids. Given such an order, then when one thinks of data fields on the input and output grids as column vectors with respect to this ordering, the transform is given by a matrix. Moreover, the locality property means that the matrix is *sparse*, i.e., that most of its entries are zero. Perhaps the simplest way to represent such matrices for our purposes is the so-called *skyline-transpose representation*, which may be described as follows:

> Let ***`NCOLS`*** and ***`NROWS`*** be the numbers of input and output grid cells, thought of as subscripts for the input and output grid, when we consider them as column vectors) and ***`NCOEF`*** the number of nonzero coefficients in the matrix.
>
> Let integer array ***`N(NROWS)`*** contain the number of nonzero coefficients in the matrix rows (i.e., with the column-vector subscripting of the

output grid).

Let integer array ***I(NCOEF)*** and real array ***C(NCOEF)*** contain respectively the input-grid column-vector subscripts and the corresponding coefficients, listed consecutively by rows (listing all the subscripts and coefficients for output grid cell # 1 first, followed by those for cell # 2, etc.)

Fortran code for the matrix multiplication looks like:

```
REAL  Y(NROWS)  !
output grid
REAL  X(NCOLS)  !
input grid
K=0
DO N=1,NROWS
   SUM0.0
   DO M=1,N(I)
      K=K+1
      SUM=SUM+C(K
      )*X(I(K))
   END DO
   X(N)=SUM
END DO
```

Fortunately, for grid-to-grid transformations we only are concerned with matrix-to-vector multiplications. We don't have to deal with the more complex data structures necessary for operations such as matrix inversion (although it should be noted that a minor variation on this data structure permits easy parallelization). Notice that there is some overhead for the sparse algorithm: we must maintain the extra index array I and we must do the extra subscripting operation **`X(I(K)).`** If most of the potential coefficients are zeros, however, this extra computational and storage overhead is less than the computational expense of doing the full-matrix operations. As it happens, for many of our applications less than one percent of the coefficients are non-zero, so the potential savings of matrix sparsification are quite large. Furthermore, the use of a methodology that enables one to “capture” the transforms in a time independent form, so that one does not have to re-calculate the transforms redundantly on demand is a much greater advancement: normally, calculating the coefficients is several times more expensive than using them.

Bilinear interpolation matrices are an important special case of geospatial transforms, for which there are always just four input cells per output value. It is worthwhile making these a special case, and providing software for that specific purpose, also. Note also that while bilinear interpolation is positive definite, it is not necessarily conservative.

The EDSS/Models-3 I/O API [Coats et al 1993, Coats 1993-2003] is the utility and input-output library used by SMOKE, CMAQ, MAQSIP, and various other models. It provides support for computing, storing, retrieving, and applying sparse matrix and bilinear-interpolation transforms. We consider such support to be one of the fundamental requirements for the infrastructure of cross-media integrated modeling systems The omission of such functionality from such efforts as the ESMF is a serious deficiency.

Applications

We first started using sparse matrices for geospatial transforms with the SMOKE atmospheric chemical emissions modeling system [Coats 1996], just over a decade ago.: Databases of chemical emissions for the US are maintained on a per-county basis as are the conversion factors from "inventory pollutants" (VOC, NO_x, SO_x) into "model species" (the forty or so distinct hydrocarbon, nitrogen, and sulfur compounds modeled by the atmospheric chemistry models). The atmospheric chemistry/ transport models, however, are grid-based, necessitating conversion from county-based geospatial data to grid-based. Because certain emissions sub-models (mobile source evaporative emissions, point source plume rise, and biogenic emissions) are meteorologically driven, SMOKE also uses sparse matrices to do the inverse transform, to generate source-level geospatial-domain meteorology to drive these latter sub-models.

Going from the "data-processing algorithm" approaches used by prior emissions models to the sparse matrix based techniques used in SMOKE reduced the computational time several hundredfold, and reduced the disk-space requirements ten-fold. This improvement is one of the critical pieces of technology that has made real-time numerical air quality prediction possible [McHenry et al 2000, 2001].

For hydrology models and hydrologically based land surface models, the fundamental element of geospatial decomposition is usually the *catchment basin* or the *hill-slope*, not the grid cell. We have used the sparse-matrix transform technique to couple the (very high resolution distributed) TOPLATS land surface/hydrology model efficiently with the (gridded) MM5 meteorology model. The transforms we have used for this purpose include the following:

- NEXRAD Stage IV gridded precipitation to TOPLATS
- MM5 gridded meteorology variables (temperature, moisture, winds, pressure, radiation, and precipitation) to TOPLATS
- TOPLATS surface fluxes to MM5 grid.

The use of these three transforms (as well as the "Coupling Mode" of the Models-3 I/O API [Coats 1995-2003] that supports data-driven cooperating-process model coupling) has allowed up to investigate a variety of one-way, one-and-a-half-way (where TOPLATS is driven by NEXRAD precipitation but MM5 ambient-atmosphere variables, as an LDAS), and two-way (pure forecast-mode) TOPLATS:MM5 couplings. We feel that "predictor-corrector" systems using both the one-and-a-half-way LDAS-mode and the two-way forecast-mode couplings may offer the opportunity to do a much better job of modeling MM5-subgridscale effects, leading to better forecast skill.

References

Coats, C. J., Jr., 1995-2003: The EDSS/Models-3 I/O API. *http:://www.baronams.com/products/ioapi/index.html*

Coats, C.J., Jr., 1996: High Performance Algorithms in the Sparse Matrix Operator Kernel Emissions (SMOKE) Modeling System. Ninth AMS Joint Conference on Applications of Air Pollution Meteorology with A&WMA, 584-588.

Coats, C. J. , Jr., A. Hanna, D. Hwang, and D. W.. Byun, 1993: Model Engineering Concepts for Air Quality Models in an Integrated Environmental Modeling System. Transactions, Regional Photochemical Measurement and Modeling Studies, Air and Waste Management Association, San Diego, CA. pp. 213-223.

Coats, C.J., A.Trayanov, J.N.McHenry, A.Xiu, A. Lario-Gibbs, and C.D. Peters-Lidard, 1999: An extension of the EDSS/Models-3 I/O API for coupling concurrent environmental models, with applications to air quality and hydrology. Preprints. 15th IIPS Conference, Amer. Meteor. Soc., Dallas, Tx.

Grell, G.A., J. Dudhia, and D.R. Stauffer, 1994: A description of the fifth-generation Penn State/NCAR mesoscale model (MM5). NCAR Technical Note TN-398, Boulder CO.

McHenry, J. N., N. Seaman, C. Coats, D. Stauffer, A. Gibbs, J.Vukovich, E. Hayes, andN. Wheeler., 2000: The NCSC-PSU numerical air quality prediction project: Initial evaluation, status, and prospects, Symposium on Atmospheric Chemistry in the 21st Century, American Meteorology Society, Boston, 95-102.

McHenry, J. N., C. Coats, B. Cameron, J. Vukovich, A. Trayanov, and T. Smith., 2001: High-Resolution Real-Time Ozone Forecasts for the August-September Texas AQS-2000 (Houston) Field Study: Forecast Process and Preliminary Evaluation. A Millenium Symposium on Atmospheric Chemistry: Past, Present, and Future of Atmospheric Chemistry. American Meteorological Society, Boston, 186-193.

Peters-Lidard, C.D., M.S.Zion, and E.F.Wood, 1997: A soil-vegetation-atmosphere transfer scheme for modeling spatially variable water and energy balance processes. J. Geophys. Res., 102, D4, 4303-4324.

Peters-Lidard, C.D., J.N. McHenry, C.J.Coats, A.Trayanov, S.Fine, K.Alapaty, and F.Pan, 1999: Re-thinking the atmospheric LSP problem from a hydrological perspective. Preprint Volume: 14th Conference on Hydrology, American Meteorological Society, Dallas, Texas.

Strang, Gilbert, 1988. Linear Algebra and Its Applications. New York: Harcourt Brace College Publishers, 1988.