Errata

Environmental Data Analysis with MATLAB® and Python

Third Edition

by William Menke

last updated 2024/05/27

Page 31, Software update, edapy01_07, dot product

I am rather aghast that the new versions of numpy, scipy and matplotlib have depreciated some common coding "phrases". The one relevant here is that a 1×1 array no longer counts as a scalar. This breaks a lot of the code. So I am now (May 24, 2024) coding the dot product $s = \mathbf{a} \cdot \mathbf{b}$ as

```
s=np.matmul(a.T,b); s=s[0,0];
```

and not as

```
s=np.matmul(a.T,b);
```

That is, `s` is overwritten by its first element. In cases where the dot product appears in an expression, I have broken it out into several lines. E.g. what was

```
t=a/sqrt(np.matmul(a.T,a));
```

becomes

```
asq=np.matmul(a.T,a); asq=asq[0,0];
t=a/sqrt(asq);
```

I have not (yet) made any effort to add to these errata specific places in the book where this kind of change has been made.

Page 45, Problem 1.2, type: “diumber” should be “number”

1.2. Write a script that defines a column-vector **a** of length $N = 12$ whose elements are the $\boxed{\text{diumber}}$ of days in the 12 months of the year, for a nonleap year. Create a similar colun-vector **b** for a leap year. The concatenate **a** and **b** into an $N \times M = 12 \times 2$ matrix **C**.

Page 148, typo in Eqn 5.22

$\mathbf{F}$ should be $\mathbf{H}$ as shown

$$\mathbf{F} = \begin{bmatrix} \mathbf{C}_d^{-1/2}\mathbf{G} \\ \mathbf{C}_h^{-1/2}\boxed{\mathbf{F}} \end{bmatrix} \quad \text{and} \quad \mathbf{f}^{obs} = \begin{bmatrix} \mathbf{C}_d^{-1/2}\mathbf{d}^{obs} \\ \mathbf{C}_h^{-1/2}\mathbf{h}^{pri} \end{bmatrix} \tag{5.22}$$

$\mathbf{H}$

Page 160, Software update, eda05_07, `las.bicg()`

I am rather aghast that the new versions of numpy, scipy and matplotlib have depreciated some common coding methods. The one relevant here is that the `tol` keyword of `la.bicg()` has been changed to `rtol`. This breaks a lot of the code.

Thus the new command

```
q=la.bicg(LO,FTh,rtol=tol,maxiter=maxit);
```

replaces the old

```
q=la.bicg(LO,FTh,tol=tol,maxiter=maxit);
```

Also, I am now explicitly casting (converting) the result to a `float`, with the new command

```
mest=gda_cvec(q[0].astype(float));
```

which replaces

```
mest=gda_cvec(q[0]);
```

However, I'm not sure this is necessary. But it doesn't hurt.

I have not (yet) made any effort to add to these errata specific places in the book where this kind of change has been made.

Page 328, Equation 10.14, error in equation

Equation 10.14 for the posterior covariance, $\mathbf{C}_{m^{est}}^{(tt)} = \sigma_d^2 \mathbf{C}_m^{(tc)} \mathbf{Q}^{-2}\ \mathbf{C}_m^{(ct)}$, is incorrect, because it does not include the effect of the covariance, $\mathbf{C}_m$, of the prior mean, $\langle \mathbf{m} \rangle$, The correct formula is

$$\mathbf{C}_{m^{est}}^{(tt)} = \mathbf{C}_m^{(tt)} - \ \mathbf{C}_m^{(tc)} \mathbf{Q}^{-1}\ \mathbf{C}_m^{(ct)} \text{ with } \mathbf{Q} \equiv \mathbf{C}_m^{(tc)} + \sigma_d^2 \mathbf{I}$$

However, like many other issues in data analysis, whether adding in the uncertainty of prior information is desirable depends upon the perspective; that is, what scatter one wants o characterize.

One possibility is that the observations really are of a random process and one has confidence that one knows the mean, $\langle \mathbf{m} \rangle$, and covariance, $\mathbf{C}_m$, of that process. One imagines that one has many repeat datasets, each of a *different* realization of the random process, and wants to know the scatter in the GPR reconstructions. Consider a target point, t, that far from any observation. The value there is fluctuating between realizations, and since no data constrain it, its scatter is controlled by $\mathbf{C}_m$, only. In this case the corrected formula should be used, for accounts for this fluctuation. Were the effect not included, then far from observations the posterior covariance would tend to zero, which would seem to violate common sense.

However, suppose that one had many repeat datasets of a deterministic process, each with the same measurement variance σ_d^2 and suppose one performed GPR on each, primarily as a tool for smoothing the data and filling in data gaps. One might be inclined to use the same choice of $\langle \mathbf{m} \rangle$ for every dataset (for example, $\langle \mathbf{m} \rangle = \mathbf{0}$). Then, the scatter among this ensemble of solutions would indeed be given by Equation 10.14. In order for the ensemble to have scatter influenced by $[\text{cov}_A\ \mathbf{m}]$, one would have to add a different realization of noise $\mathbf{n}$ to $\langle \mathbf{m} \rangle$ (for example $\langle \mathbf{m} \rangle = \mathbf{0} + \mathbf{n}$) during the GPR solution process (with $\mathbf{n}$ having covariance, $\mathbf{C}_m$). In practice, this is rarely (if ever) done; one sticks with the same $\langle \mathbf{m} \rangle$.

This error has been corrected in the exemplary code and solutions as of May 28, 2024.