Problem 1: A closer look at Broyden's Method

Problem Statement

Jacobian surrogates in Broyden's method are defined as:

$$J^k := J^{k-1} + \frac{1}{\|\Delta x^k\|_2^2}\left[\Delta F^k - J^{k-1}\Delta x^k\right]\left[\Delta x^k\right]^T$$

where

$$\Delta F^k := F(x^k) - F(x^{k-1}),\ \Delta x^k := x^k - x^{k-1}$$

a

For $A \in \mathbb{R}^{n\times n}$ and $u, v \in \mathbb{R}^n$, the Sherman-Morrison formula states:

$$\left(A + uv^T\right)^{-1} = A^{-1} - \frac{A^{-1}uv^TA^{-1}}{1 + v^TA^{-1}u}$$

Using this, prove that:

$$B^k := \left[J^k\right]^{-1} = B^{k-1} + \frac{\Delta x^k - B^{k-1}\Delta F^k}{\left[\Delta x^k\right]^T B^{k-1}\Delta F^k}\left[\Delta x^k\right]^T B^{k-1}$$

Comment why this recurrence relation in B^k is useful for solving the quasi-Newton method.

b

Recall that Broyden's idea iteratively solves

$$\min_{J^k\in\mathbb{R}^{n\times n}} \|J^k - J^{k-1}\|_F$$

subject to $J^k\Delta x^k = \Delta F^k$ to define the sequence of Jacobian surrogates. Another method proposed calculates the surrogate Jacobian inverses by solving:

$$\min_{J^k\in\mathbb{R}^{n\times n}} \|C^k - C^{k-1}\|_F$$

subject to $C^k\Delta F^k = \Delta x^k$. This is the so-called "bad method". Compute C_k in terms of C^{k-1}, Δx^k, and ΔF^k. Comment if you think it's really that bad! Hint: exploit the similarity between the two functions.

Solution

a

For convenience in typing, we'll drop the indices and deltas for now. Also,for brevity, let $\sigma = \frac{1}{\|x\|_2^2}$. Let:

$$A := J, u := F - Jx,\ v := \sigma x$$
$$B = A^{-1} = J^{-1}$$

Substituting:

$$(J + \sigma(F - Jx)x^T)^{-1} = B - \frac{B(F - Jx)x^TB}{1 + \sigma x^TB(F - Jx)}$$

$$= B - \frac{(BF - BJx)x^T B}{1 + \sigma x^T BF - \sigma x^T BJx}$$

$$B \coloneqq J^{-1} \implies BJ = JB = I$$

$$\implies = B - \frac{(BF - x)\sigma x^T B}{1 + \sigma x^T BF - \sigma x^T x} = B + \frac{(x - BF)\sigma x^T B}{1 + \sigma x^T BF - \sigma x^T x}$$

$$= B + \frac{(x - BF)\sigma x^T B}{1 + \sigma x^T BF - \sigma x^T x}$$

$$\|x\|_2^2 = x^T x \implies \sigma x^T x = 1$$

$$\implies = B + \frac{(x - BF)\sigma x^T B}{\sigma x^T BF} = B + \frac{x - BF}{x^T BF} x^T B$$

Reinserting indices and deltas:

$$B^k = B^{k-1} + \frac{\Delta x^k - B^{k-1}\Delta F^k}{[\Delta x^k]^T B^{k-1} F}[\Delta x^k]^T B$$

□

This relationship is useful because it removes the requirement to invert the matrix J^k and instead calculates the inverse directly, significantly reducing computational overhead.

b

We will follow much the same procedure as outlined in the notes; however, we will replace J with C and $\Delta x \iff \delta F$. (My gut says it will result in the same functional form, with only the above variables replaced as indicated...)

Define $\Delta C \coloneqq C^k - Ck - 1$. We seek to solve

$$\min \|\Delta C\|_F$$

subject to

$$(C^{k-1} + \Delta C)\Delta F^k = \Delta x^k$$

Note, too, that because $\|\cdot\|_F \geq 0$, the minimizer of this function is the same as $\|\cdot\|_F^2$. From here we can rearrange the constraint function and define β:

$$\Delta C \Delta F^k = \Delta x^k - J^{k-1}\Delta F^k, \ \beta \coloneqq \Delta x^k - J^{k-1}\Delta F^k$$

$$\Delta C = \begin{pmatrix} z_1^T \\ \vdots \\ z_n^T \end{pmatrix}$$

i.e. z_i^T is the i-th row of ΔC. So:

$$\min \|\delta C\|_F^2$$

subject to:

$$\Delta C \Delta F^k = \beta$$

$$\equiv \min \sum_{i=1}^{n} \|z_i\|_2^2$$

subject to:

$$\left[\Delta F^k\right]^T z_i = \beta_i,\ i \in 1, \ldots, n$$

which is solved by

$$z_i^* = \frac{\beta}{\|\Delta F^k\|_2^2} \Delta F^k$$

Optimal ΔC is thus given by

$$\begin{pmatrix} [z_1^*]^T \\ \vdots \\ [z_n^*]^T \end{pmatrix} = \begin{pmatrix} \frac{1}{\|\Delta F^k\|_2^2} \beta_1 \left[\Delta F^k\right]^T \\ \vdots \\ \frac{1}{\|\Delta F^k\|_2^2} \beta_n \left[\Delta F^k\right]^T \end{pmatrix} = \frac{1}{\|\Delta F^k\|_2^2} \left[\Delta x^k - C^{k-1} \Delta F^k\right] \left[\Delta F^k\right]^T$$

Thus:

$$C^k = \frac{\Delta x^k - C^{k-1} \Delta F^k}{\|\Delta F^k\|^2} \left[\Delta F^k\right]^T$$

Shock beyond shock! Swapping Δx and ΔF in the original equation results in them being swapped in the resulting equation!

As for the relative badness – given its name, and Broyden's comment in his original paper that "...this method appears in practice to be unsatisfactory, it will be discussed no further at this stage." [1] I would suspect it likely displays inferior numerical properties. Without any rigor, I would imagine constraining the updates of the inverse Jacobian surrogate to be minimal in the Frobenious norm would be a worse approximation than that constraint upon the Jacobian itself and using the Sherman-Morrison relationship to invert it. In essence, a minimal change in the inverse is likely less"physically correct".

[1] See page 582 of "A class of methods for solving nonlinear simultaneous equations" by Broyden, available here (click me!) as a PDF.

Problem 2: Solving 'em power flows

Problem Statement

Use the provided Matlab files (or write your own) to simulate the 3-bus system discussed in class. Use $\epsilon = 10^{-10}$ for the terminating condition. Make sure the code converges to:

$$\begin{pmatrix} \theta_2^* \\ \theta_3^* \\ v_3^* \end{pmatrix} = \begin{pmatrix} -0.0101 \\ -0.0635 \\ 0.9816 \end{pmatrix}$$

a

Discuss what you think would constitute a physically meaningful solution to the power flow equations. Qualitatively discuss how you would enforce a solver to produce such a meaningful solution

b

By varying the starting point for the actual NR iteration, explore if the solution provided above is the unique solution to the power flow problem for this example. If not, comment on whether this other solution you obtained is physically meaningful.

Solution

a

The final output of the files as distributed (and ran in Octave cuz I didn't want to install Matlab) are:

```
The last iterate [t2 t3 v3] = [-0.01009   -0.063514     0.98159]
```

which do indeed match the expected values.

"Physically meaningfull" would involve, primarily, positive voltages. The are per-unit, so "negative" voltage magnitude is meaningless and indicates a bad solution. Furthermore, any wildly unxpected value for voltage magnitude (i.e. "far" from 1) would be subject to scrutiny, as either the solver has converged to a non-realizable solution...or things are about to go very, very wrong on the system.

Similarly, I would suspect that phase angles far from 0 (i.e close to $2n\pi + 1$) would be suspect, unless something about the system is known to structurally cause large phase shifts (If I recall correctly, the WEC can show relatively large phase shifts given its "trunk and branch" topology, compared to the eastern interconnect which is more of a mesh).

Enforcement of this would involve incorporating boundaries of exected values and requiring the solver to restart from a perturbed initial position until converges to an acceptable solution. For instance, most solvers start from a previously known solution. If this fails, perhaps a "flat start" initial condition may produce a more realistic result.

Also, for shame, using inv() directly in the code, tsk tsk!

b

It certainly is not. Setting the initial conditions to:

```
t2=1; t3=-1; v3=2;
```

The solver produces a result of:

```
The last iterate [t2 t3 v3] = [-6.57205     -23.1512     -0.06764]
```

which has an error of only $6.2701 * 10^{-13}$. But, as mentiond before, this negative P.U. voltage is a sure sign something is awry.