You Could Have Invented Spectral Sequences

Timothy Y. Chow

Introduction

The subject of spectral sequences has a reputation for being difficult for the beginner. Even G. W. Whitehead (quoted in John McCleary [4]) once remarked, "The machinery of spectral sequences, stemming from the algebraic work of Lyndon and Koszul, seemed complicated and obscure to many topologists."

Why is this? David Eisenbud [1] suggests an explanation: "The subject of spectral sequences is elementary, but the notion of the spectral sequence of a double complex involves so many objects and indices that it seems at first repulsive." I have heard others make similar complaints about the proliferation of subscripts and superscripts.

My own explanation, however, is that spectral sequences are often not taught in a way that explains how one might have come up with the definition in the first place. For example, John McCleary's excellent text [4] says, "The user, however, needs to get acquainted with the manipulation of these gadgets without the formidable issue of their origins." Without an understanding of where spectral sequences come from, one naturally finds them mysterious. Conversely, if one *does* see where they come from, the notation should not be a stumbling block.

Fools rush in where angels fear to tread, so my goal below is to make you, the reader, feel that you could have invented spectral sequences (on a very good day, to be sure!). I assume familiarity with homology groups, but little more. Everything here is known to the cognoscenti, but my hope is to make the ideas accessible to more than the lucky few who are able to have the right conversation with the right expert at the right time.

Readers who are interested in the history of spectral sequences and how they were in fact invented should read [3], which gives a definitive account.

Simplifying Assumptions

Throughout, we work over a field. All chain groups are finite-dimensional, and all filtrations (explained below) have only finitely many levels. In the "real world", these assumptions may fail, but the essential ideas are easier to grasp in this simpler context.

Graded Complexes

Chain complexes that occur "in nature" often come with extra structure in addition to the boundary map. Certain kinds of extra structure are particularly common, so it makes sense to find a systematic method for exploiting such features. Then we do not have to reinvent the wheel each time we want to compute a homology group.

Timothy Y. Chow is a member of the research staff at the Center for Communications Research in Princeton, New Jersey. His email address is `tchow@alum.mit.edu`.

Here is a simple example. Suppose we have a chain complex

$$\cdots \xrightarrow{\partial} C_{d+1} \xrightarrow{\partial} C_d \xrightarrow{\partial} C_{d-1} \xrightarrow{\partial} \cdots$$

that is "graded", i.e., each C_d splits into a direct sum

$$C_d = \bigoplus_{p=1}^{n} C_{d,p}$$

and moreover the boundary map ∂ respects the grading in the sense that $\partial C_{d,p} \subseteq C_{d-1,p}$ for all d and p. Then the grading allows us to break up the computation of the homology into smaller pieces: simply compute the homology in each grade independently and then sum them all up to obtain the homology of the original complex.

Unfortunately, in practice we are not always so lucky as to have a grading on our complex. What we frequently have instead is a *filtered complex,* i.e., each C_d comes equipped with a nested sequence of submodules

$$0 = C_{d,0} \subseteq C_{d,1} \subseteq C_{d,2} \subseteq \cdots \subseteq C_{d,n} = C_d$$

and the boundary map respects the filtration in the sense that

$$\partial C_{d,p} \subseteq C_{d-1,p} \tag{1}$$

for all d and p. (Note: The index p is called the *filtration degree.* Here it has a natural meaning only if $0 \le p \le n$, but throughout this paper, we sometimes allow indices to "go out of bounds," with the understanding that the objects in question are zero in that case. For example, $C_{d,-1} = 0$.)

Although a filtered complex is not quite the same as a graded complex, it is similar enough that we might wonder if a similar "divide and conquer" strategy works here. For example, is there a natural way to break up the homology groups of a filtered complex into a direct sum? The answer turns out to be yes, but the situation is surprisingly complicated. As we shall now see, the analysis leads directly to the concept of a spectral sequence.

Let us begin by trying naïvely to "reduce" this problem to the previously solved problem of graded complexes. To do this we need to express each C_d as a direct sum. Now, C_d is certainly not a direct sum of the $C_{d,p}$; indeed, $C_{d,n}$ is already all of C_d. However, because C_d is a finite-dimensional vector space (recall the assumptions we made at the outset), we can obtain a space isomorphic to C_d by modding out by any subspace U and then direct summing with U; that is to say, $C_d \simeq (C_d/U) \oplus U$. In particular, we can take $U = C_{d,n-1}$. Then we can iterate this process to break U itself down into a direct sum, and continue all the way down. More formally, define

$$E^0_{d,p} \overset{\text{def}}{=} C_{d,p}/C_{d,p-1} \tag{2}$$

for all d and p. (*Warning:* There exist different indexing conventions for spectral sequences; most authors write $E^0_{p,q}$ where $q = d - p$ is called the *complementary degree.* The indexing convention I use here is the one that I feel is clearest pedagogically.) Then

$$C_d \simeq \bigoplus_{p=1}^{n} E^0_{d,p}. \tag{3}$$

The nice thing about this direct sum decomposition is that the boundary map ∂ naturally induces a map

$$\partial^0 : \bigoplus_{p=1}^{n} E^0_{d,p} \to \bigoplus_{p=1}^{n} E^0_{d-1,p}$$

such that $\partial^0 E^0_{d,p} \subseteq E^0_{d-1,p}$ for all d and p. The reason is that two elements of $C_{d,p}$ that differ by an element of $C_{d,p-1}$ get mapped to elements of $C_{d-1,p}$ that differ by an element of $\partial C_{d,p-1} \subseteq C_{d-1,p-1}$, by equation (1).

Therefore we obtain a graded complex that splits up into n pieces:

$$\begin{array}{ccccccccc}
\cdots & \xrightarrow{\partial^0} & E^0_{d+1,n} & \xrightarrow{\partial^0} & E^0_{d,n} & \xrightarrow{\partial^0} & E^0_{d-1,n} & \xrightarrow{\partial^0} & \cdots \\
\cdots & \xrightarrow{\partial^0} & E^0_{d+1,n-1} & \xrightarrow{\partial^0} & E^0_{d,n-1} & \xrightarrow{\partial^0} & E^0_{d-1,n-1} & \xrightarrow{\partial^0} & \cdots \\
 & & \vdots & & \vdots & & \vdots & & \\
\cdots & \xrightarrow{\partial^0} & E^0_{d+1,1} & \xrightarrow{\partial^0} & E^0_{d,1} & \xrightarrow{\partial^0} & E^0_{d-1,1} & \xrightarrow{\partial^0} & \cdots
\end{array} \tag{4}$$

Now let us define $E^1_{d,p}$ to be the pth graded piece of the homology of this complex:

$$E^1_{d,p} \overset{\text{def}}{=} H_d(E^0_{d,p}) = \frac{\ker \partial^0 : E^0_{d,p} \to E^0_{d-1,p}}{\operatorname{im} \partial^0 : E^0_{d+1,p} \to E^0_{d,p}} \tag{5}$$

(For those comfortable with relative homology, note that $E^1_{d,p}$ is just the relative homology group $H_d(C_p, C_{p-1})$.) Still thinking naïvely, we might hope that

$$\bigoplus_{p=1}^{n} E^1_{d,p} \tag{6}$$

is the homology of our original complex. Unfortunately, this is too simple to be true. Although *each term* in the the complex $(\bigoplus_p E^0_{d,p}, \partial^0)$—known as the *associated graded complex* of our original filtered complex (C_d, ∂)—is isomorphic to the corresponding term in our original complex, this does not guarantee that the two complexes will be isomorphic *as chain complexes.* So although $\bigoplus_p E^1_{d,p}$ does indeed give the homology of the associated graded complex, it may *not* give the homology of the original complex.

Analyzing the Discrepancy

This is a little disappointing, but let's not give up just yet. The associated graded complex is so closely related to the original complex that even if its homology isn't exactly what we want, it ought to be a reasonably good approximation. Let's carefully examine the discrepancy to see if we can fix the problem.

Moreover, to keep things as simple as possible, let us begin by considering the case $n = 2$. Then the array in diagram (4) has only two levels, which we shall call the "upstairs" ($p = 2$) and "downstairs" ($p = 1$) levels.

The homology group H_d that we really want is Z_d/B_d, where Z_d is the space of cycles in C_d and B_d is the space of boundaries in C_d. Since C_d is filtered, there is also a natural filtration on Z_d and B_d:

$$0 = Z_{d,0} \subseteq Z_{d,1} \subseteq Z_{d,2} = Z_d$$

and

$$0 = B_{d,0} \subseteq B_{d,1} \subseteq B_{d,2} = B_d.$$

Recall that we have been trying to find a natural way of decomposing H_d into a direct sum. Just as we observed before that C_d is not a direct sum of $C_{d,1}$ and $C_{d,2}$, we observe now that Z_d/B_d is not a direct sum of $Z_{d,1}/B_{d,1}$ and $Z_{d,2}/B_{d,2}$; indeed, $Z_{d,2}/B_{d,2}$ by itself is already the entire homology group. But again we can use the same trick of modding out by the "downstairs part" and then direct summing with the "downstairs part" itself:

$$\frac{Z_d}{B_d} \simeq \frac{Z_d + C_{d,1}}{B_d + C_{d,1}} \oplus \frac{Z_d \cap C_{d,1}}{B_d \cap C_{d,1}} = \frac{Z_{d,2} + C_{d,1}}{B_{d,2} + C_{d,1}} \oplus \frac{Z_{d,1}}{B_{d,1}}.$$

Now, the naïve hope would be that

$$E^1_{d,2} \stackrel{?}{\simeq} \frac{Z_{d,2} + C_{d,1}}{B_{d,2} + C_{d,1}} \tag{7}$$

and

$$E^1_{d,1} \stackrel{?}{\simeq} \frac{Z_{d,1}}{B_{d,1}}, \tag{8}$$

and even that the numerators and denominators in equations (7) and (8) are precisely the "cycles" and "boundaries" in the definition (equation (5) above) of $E^1_{d,p}$. For then expression (6) would give us a direct sum decomposition of H_d. Unfortunately, in general, neither (7) nor (8) holds. Corrections are needed.

Let us first look "downstairs" at $E^1_{d,1}$. The "cycles" of $E^1_{d,1}$ are the cycles in $E^0_{d,1}$, and the "boundaries" are the image I of the map

$$\partial^0 : E^0_{d+1,1} \to E^0_{d,1}.$$

The space of cycles in $E^0_{d,1}$ is $Z_{d,1}$, which is the numerator in equation (8). However, the image I is *not* $B_{d,1}$, for $B_{d,1}$ is the part of B_d that lies in $C_{d,1}$, and while this *contains* I, it may also contain other things. Specifically, the map ∂ may carry some elements $x \in C_{d+1}$ down from "upstairs" to "downstairs," whereas I only captures boundaries of elements that were already downstairs to begin with. Therefore, $Z_{d,1}/B_{d,1}$ is a *quotient* of $E^1_{d,1}$.

Now let us look "upstairs" at $E^1_{d,2}$. In this case, the space of "boundaries" of $E^1_{d,2}$ is $B_{d,2} + C_{d,1}$, which is the denominator in equation (7). However, the space of "cycles" in this case is the kernel K of the map

$$\partial^0 : E^0_{d,2} \to E^0_{d-1,2},$$

which, by definition of E^0, is the map

$$\partial^0 : \frac{C_{d,2}}{C_{d,1}} \to \frac{C_{d-1,2}}{C_{d-1,1}}.$$

Thus we see that K contains not only chains that ∂ sends to zero but also any chains that ∂ sends "downstairs" to $C_{d-1,1}$. In contrast, the elements of $Z_{d,2} + C_{d,1}$ are more special: their boundaries are boundaries of chains that come from $C_{d,1}$. Hence

$$\frac{Z_{d,2} + C_{d,1}}{B_{d,2} + C_{d,1}}$$

is a *subspace* of $E^1_{d,2}$, the subspace of elements whose boundaries are boundaries of $C_{d,1}$-chains.

Intuitively, the problem is that the associated graded complex only "sees" activity that is confined to a single horizontal level; everything above and below that level is chopped off. But in the original complex, the boundary map ∂ may carry things down one or more levels (it cannot carry things *up* one or more levels because ∂ respects the filtration), and one must therefore correct for this inter-level activity.

The Emergence of Spectral Sequences

The beautiful fact that makes the machinery of spectral sequences work is that both of the above corrections to the homology groups $E^1_{d,p}$ can be regarded as "homology groups of homology groups"!

Notice that ∂ induces a natural map—let us call it ∂^1—from $E^1_{d+1,2}$ to $E^1_{d,1}$, for all d, for the boundary of any element in $E^1_{d+1,2}$ is a cycle that lies in $C_{d,1}$, and thus it defines an element of $E^1_{d,1}$. The key claims (for $n = 2$) are the following.

- *Claim 1. If we take $E^1_{d,1}$ and mod out by the image of ∂^1, then we obtain $Z_{d,1}/B_{d,1}$.* To see this, just check that the image of ∂^1 gives all the boundaries that lie in $C_{d,1}$.
- *Claim 2. The kernel of ∂^1 is a subspace of $E^1_{d,2}$ isomorphic to*

$$\frac{Z_{d+1,2} + C_{d+1,1}}{B_{d+1,2} + C_{d+1,1}}.$$

Again, simply check that the kernel consists just of those elements whose boundary equals a boundary of some element of $C_{d+1,1}$.

We can visualize these claims by drawing the following diagram.

$$
\begin{array}{ccccccccc}
0 & & 0 & & 0 & & 0 & & 0 \\
& \searrow\partial^1 & & \searrow\partial^1 & & \searrow\partial^1 & & \searrow\partial^1 & \\
\cdots & & E^1_{d+1,2} & & E^1_{d,2} & & E^1_{d-1,2} & & \cdots \\
& \searrow\partial^1 & & \searrow\partial^1 & & \searrow\partial^1 & & \searrow\partial^1 & \\
\cdots & & E^1_{d+1,1} & & E^1_{d,1} & & E^1_{d-1,1} & & \cdots \\
& \searrow\partial^1 & & \searrow\partial^1 & & \searrow\partial^1 & & \searrow\partial^1 & \\
0 & & 0 & & 0 & & 0 & & 0
\end{array}
\tag{9}
$$

Diagram (9) is a collection of chain complexes; it's just that the chain complexes do not run horizontally as in diagram (4), but slant downwards at a $45°$ angle, and each complex has just two nonzero terms. (*Reminder:* Our indexing convention is different from that of most authors, whose diagrams will therefore look "skewed" relative to diagram (9).) If we now define $E^2_{d,p}$ to be the homology, i.e.,

$$
E^2_{d,p} \stackrel{\text{def}}{=} H_d(E^1_{d,p}) = \frac{\ker \partial^1 : E^1_{d,p} \to E^1_{d-1,p-1}}{\operatorname{im} \partial^1 : E^1_{d+1,p+1} \to E^1_{d,p}}, \tag{10}
$$

then the content of Claim 1 and Claim 2 is that $E^2_{d,1} \oplus E^2_{d,2}$ is (finally!) the correct homology of our original filtered complex.

For the case $n = 2$, this completes the story. The sequence of terms E^0, E^1, E^2 is the spectral sequence of our filtered complex when $n = 2$. We may regard E^1 as giving a first-order approximation of the desired homology, and E^2 as giving a second-order approximation—which, when $n = 2$, is not just an approximation but the true answer.

What if $n > 2$? The definitions (2), (5), and (10) still make sense, but now E^2 will not in general give the true homology, because E^2 only takes into account interactions between *adjacent* levels in diagram (4), but ∂ can potentially carry things down two or more levels. Therefore we need to consider further terms $E^3, E^4, \ldots, E^n$. For example, to define E^3, we can check that ∂ induces a natural map—call it ∂^2—from $E^2_{d+1,p+2}$ to $E^2_{d,p}$, for all d and p. One obtains a diagram similar to diagram (9), except with (E^2, ∂^2) instead of (E^1, ∂^1), and with each arrow going down *two* levels instead of one. Then $E^3_{d,p}$ is $(\ker \partial^2)/(\operatorname{im} \partial^2)$ at $E^2_{d,p}$. In general, the picture for E^r has arrows labeled ∂^r dropping down r levels from $E^r_{d+1,p+r}$ to $E^r_{d,p}$, and E^{r+1} is defined to be the homology of (E^r, ∂^r).

The verification that, for general n, $H_d \simeq \bigoplus_p E^n_{d,p}$ is a conceptually straightforward generalization of the ideas we have already seen, but it is tedious so we omit the details.

What Good Is All This?

In analysis, the value of having a series approximation converging to a quantity of interest is familiar to every mathematician. Such an approximation is particularly valuable when just the first couple of terms already capture most of the information.

Similar remarks apply to spectral sequences. One common phenomenon is for a large number of the $E^r_{d,p}$ and/or the boundary maps ∂^r to become zero for small values of r. This causes the spectral sequence to stabilize or *collapse* rapidly, allowing the homology to be computed relatively easily. We illustrate this by sketching the proof of Theorem 2 in a paper of Phil Hanlon [2]. This is far from a "mainstream" application of spectral sequences, but it has the great advantage of requiring very little background knowledge to follow. Readers who know enough topology may wish instead to proceed directly to the standard examples that may be found in any number of textbooks.

Let Q be a finite partially ordered set that is *ranked*—i.e., every maximal totally ordered subset has the same number of elements, so that every element can be assigned a rank (namely, a natural number indicating its position in any maximal totally ordered subset containing it)—and that is equipped with an order-reversing involution $x \mapsto x^*$. Let

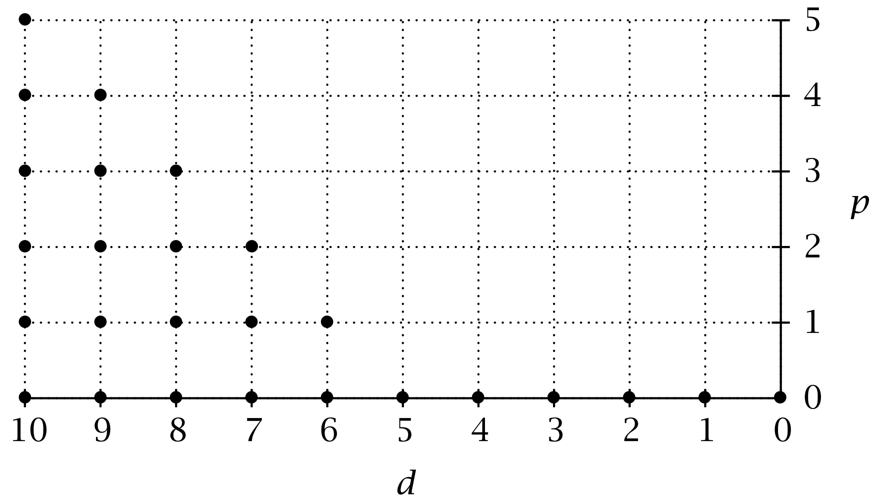
$$
\gamma = \{\alpha_1, \alpha_2, \ldots, \alpha_t\} \quad \text{where} \quad \alpha_1 < \alpha_2 < \cdots < \alpha_t \tag{11}
$$

be a totally ordered subset of Q. We say that γ is *isotropic* if $\alpha_i \neq \alpha_j^*$ for all i and j.

Now adjoin a minimum element $\hat{0}$ and a maximum element $\hat{1}$ to Q, and consider the family of all totally ordered subsets of the resulting partially ordered set. These form an abstract simplicial complex Δ, and we can consider its simplicial homology groups H_d. We can also restrict attention to the isotropic totally ordered subsets; these form a subcomplex Δ^0, which has its own homology groups H^0_d.

Hanlon's Theorem 2 says that if Q is Cohen-Macaulay and its maximal totally ordered subsets have m elements, then $H^0_d = 0$ if $0 \le d < m/2$. The definition of *Cohen-Macaulay* need not concern us here; it suffices to know that Cohen-Macaulay partially ordered sets satisfy a certain homological property (given in Hanlon's paper). In particular, knowing that Q is Cohen-Macaulay gives us information about H_d.

In order to deduce something about H_d^0 from the information we have about H_d, we seek a relationship between H_d and H_d^0. Given γ as in equation (11), Hanlon's key idea is to let $\rho(\gamma)$ be the rank of α_i, where i is maximal subject to the condition that $\alpha_i^* = \alpha_j$ for some $j > i$. Then $\rho(\gamma) = 0$ if and only if γ is isotropic, but more importantly, applying the boundary map can clearly only decrease ρ, so ρ induces a filtration on Δ. Specifically, we obtain the pth level of the filtration by restricting to those γ such that $\rho(\gamma) \le p$. Therefore we obtain a spectral sequence! This gives a relationship between $H_d^0 = E_{d,0}^1$ and the limit H_d of the spectral sequence.



The heart of Hanlon's proof is to analyze E^1. He shows that $E_{d,p}^1 = 0$ except possibly for certain pairs (d, p). For instance, when $m = 10$, $E_{d,p}^1 = 0$ except possibly for the pairs (d, p) marked by dots in the diagram to the right.

If you imagine the 45° boundary maps, then you can see that some potentially complicated things may be happening for $d \ge 5 = m/2$, but for $m/2 > d$, $E_{d,p}^2$ will be isomorphic to $E_{d,p}^1$ for all p. In fact, $E_{d,p}^r \simeq E_{d,p}^1$ for all $r \ge 1$ when $m/2 > d$; the boundary maps slant more and more as r increases, but this makes no difference. Therefore just by computing E^1, we have computed the full homology group for certain values of d. In particular, $H_d^0 \simeq H_d$ for $m/2 > d$. It turns out that the Cohen-Macaulay condition easily implies that $H_d = 0$ for $m/2 > d$, so this completes the proof.

A Glimpse Beyond

When our simplifying assumptions are dropped, a lot of complications can arise. Over an arbitrary commutative ring, equation (3) need not hold; not every short exact sequence splits, so there may be extension issues. When our finiteness conditions are relaxed, one may need to consider E^r for arbitrarily large r, and the spectral sequence may not converge. Even if it does converge, it may not converge to the desired homology. So in many applications, life is not as easy as it may have seemed from the above discussion; nevertheless, our simplified setting can still be thought of as the "ideal" situation, of which more realistic situations are perturbations.

We should also mention that spectral sequences turn out to be such natural gadgets that they arise not only from filtered complexes, but also from double complexes, exact couples, etc. We cannot even begin to explore all these ramifications here, but hope that our tutorial will help you tackle the textbook treatments with more confidence.

Why the Adjective "Spectral"?

A question that often comes up is where the term "spectral" comes from. The adjective is due to Leray, but he apparently never published an explanation of why he chose the word. John McCleary (personal communication) and others have speculated that since Leray was an analyst, he may have viewed the data in each term of a spectral sequence as playing a role that the eigenvalues, revealed one at a time, have for an operator. If any reader has better information, I would be glad to hear it.

Acknowledgments

This paper started life as some lecture notes accompanying a talk I gave while I was a general member of the Mathematical Sciences Research Institute in the spring of 1997, supported in part by a National Science Foundation postdoctoral fellowship. Many thanks go to Art Duval, who went through the notes with a fine-toothed comb and provided detailed suggestions for improving the exposition, as well as repeatedly encouraging me to publish a proper writeup. I thank John McCleary for pointing out references [3] and [5] to me; the latter is another exposition of spectral sequences that is targeted at the beginner. Finally, I thank Phil Hanlon for helping me understand spectral sequences in the first place. He deserves at least as much credit as I do for any merits of the approach expounded in this paper.

References

[1] David Eisenbud, *Commutative Algebra With a View Toward Algebraic Geometry,* Springer-Verlag, 1995.

[2] Phil Hanlon, A note on the homology of signed posets, *J. Algebraic Combin.* **5** (1996), 245–250.

[3] John McCleary, A history of spectral sequences: Origins to 1953, in *History of Topology,* edited by Ioan M. James, North Holland, 1999, 631–663.

[4] ____ , *A User's Guide to Spectral Sequences,* 2nd ed., Cambridge University Press, 2001.

[5] Barry Mitchell, Spectral sequences for the layman, *Amer. Math. Monthly* **76** (1969), 599–605.