Lecture 7 Script

Stability of Graph Filters to Relative Perturbations

Foreword

1. In this lecture, we study the stability of graph filters to relative perturbations. This is work by Alejandro Ribeiro and Fernando Gama.

2. We consider relative perturbations of shift operators such that the difference between the shift operator S and its shifted version S hat is a symmetric additive term of the form E times S plus S times E.

3. The norm of the error matrix in (1) is a measure of how close, S hat and S are. We have seen that graphs that are permutations of each other are equivalent from the perspective of running graph filters. Thus, a more convenient perturbation model is to consider a relationship in which we multiply the left hand side by a permutation matrix P sub zero.

4. We measure the size of the perturbation with the norm of this other error matrix E. We can write a relationship of the form in (2) for any permutation matrix P_0.

5. Naturally, we want to consider the permutation for which the norm of the error is minimized. The bounds we will derive here apply to any pair of shift operators that are related as per (2). They will be tightest however, for the permutation matrix P sub zero for which the nom of the error E is the smallest.

Properties of the Perturbation

1. Let's study some properties of the perturbation.

2. There are two aspects of the perturbation matrix E in (2) that are important in seizing its effect on a graph filter. The norm of E and the difference between the eigenvectors of S and E. As a shorthand for the norm of E we define the constant epsilon.

3. To measure the difference between the eigenvectors of S and E, we consider the eigenvector decomposition of the shift operator to write S as V times Lambda times

V-Hermitian and the eigenvector decomposition of the error matrix to write E as U times M times U Hermitian. We then define the eigenvector misalignment constant delta, which involves the norm of the difference between U and V, plus one, squared, minus one.

4. If the eigenvectors of S and its perturbation E are the same, we have that U equals V and that, consequently, delta equals zero. As the eigenvectors grow more dissimilar, the misalignment constant grows.

5. An important ancillary remark is that since the matrices V and U are unitary, their norms are at most one. Thus, the constant delta cannot exceed eight. It is never too large. The reason for defining this constant is that it has an effect on the stability bounds we are about to derive. We want to have a concrete handle to understand the effect of perturbations when eigenvectors are known to be close to each other.

Integral Lipschitz Filters

1. We further introduce the notion of integral Lipschitz filters.

2. We will see that we cannot allow for arbitrary filters if we are to have stability to perturbations. The restriction we impose is that our filters be integral Lipschitz. Specifically, we require that for any pair of values, lambda one and lambda two, the frequency response of the graph filter, denoted here as h of lambda, be such that: The difference between the values h of lambda two and h of lambda one be bounded by a constant C, multiplying the ratio between the difference of lambda two and lambda one and the average of lambda one and lambda two. All of these quantities within absolute values.

3. The constant C in (5) is the integral Lipschitz constant of the filter. The condition in (5) can be read as requiring the filter's frequency response to be Lipschitz in any interval Lambda one Lambda two with a Lipschitz constant that is inversely proportional to the interval's midpoint: Lambda one plus Lambda 2 divided by two.

4. To understand this condition better, recall that the filters we are working with are analytic. They are therefore, in particular, differentiable. In this case the condition in (5) implies that the derivative of the frequency response must be such that the product between lambda and the derivative of the filter's frequency response h prime of lambda be bounded by C in absolute value.

5. Thus, filters that are integral Lipschitz must have frequency responses that have to be flat for large lambda, but can vary very rapidly around lambda equal zero. We will restrict our filters to be integral Lipschitz in the remainder of this presentation.

Stability to Relative Perturbations

1. With these definitions in place, we study the stability of filters to relative perturbations.

2. In particular, we will prove the following theorem. Consider shift operators S and S hat that are related as per the perturbation model in Equation (2) along with respective filters, H of S and H of S hat. These filters are made up of the same coefficients h k instantiated on these two different shift operators S and S hat. If the filters are integral Lipschitz with constant C, the operator distance modulo permutation between the two filters is bounded by: Two. Times the Lipschitz constant of the filter C. Times one plus delta square-root-of-n. Times epsilon. Plus a term that is of order Epsilon squared. In this bound: Epsilon is the norm of the error matrix. delta is the eigenvector misalignment constant we just defined. And n is the number of nodes of the graph.

3. Theorem 1 shows that filters are Lipschitz stable with respect to relative perturbations of the graph with stability constant 2-C times one plus delta square-root-of-n. This stability constant is affected by the filter’s integral Lipschitz constant C. This is a value that is controllable through filter design. The stability is also affected by a term that involves the eigenvector misalignment constant, the factor one plus delta square-root-of-n. This term depends on the structure of the perturbations that are expected in a particular problem but it cannot be affected by judicious filter choice.

Proof of Stability Theorem

1. We begin work on the proof of the stability theorem.

2. The proof of theorem one uses the graph Fourier transform representation of graph filters. During the analysis, we will encounter the products E times v_i between the error matrix E and the eigenvectors v sub i of S. The following section introduces a Lemma that provides a characterization of these products.

Eigenvector Perturbation Lemma

1. (Empty)

2. The eigenvector perturbation Lemma, considers a given error matrix E and a given shift operator S and looks at the i-th eigen**vector** of S denoted as v sub i and the i-th eigen**value** of E denoted as m sub i. We define the matrix E sub i such that the product E times v_i is written as the sum m sub i times v sub i plus E sub i times v sub i. The Matrix E_i has a norm that can be bounded by the product epsilon times delta. In this product, epsilon is the norm of the error matrix and delta is the misalignment constant we introduced.

Proof of Lemma 1

1. The proof of the Lemma follows from some simple algebraic manipulations. Define the matrix E sub i as the difference between E and V times M times V-Hermitian so that we can write the error matrix E as the sum of V M V-Hermitian and E sub i.

2. Observe that the matrix, V M V-Hermitian has the eigenvectors of S because of V and the eigenvalues of E because of M. Since the vector v sub i is the i-th eigenvector of S, it is the i-th column of V. Thus the product V M V-Hermitian times v i is simply m sub i times v sub i. Where, we recall, m_i is the i-th eigenvalue of E.

3. We substitute this expression back in (10) to write E v_i as the sum of m_i times v i and E_i times v_i.

4. We note that (11) and (8) are the same, thus the result follows if we show that the norm of the matrix E_i in (11) is bounded by the product epsilon times delta.

5. To show that this is true, we use the eigenvector decomposition E equals, U M U-Hermitian to write the matrix E sub i as we show in (12).

6. We now consider the eigenvector matrix difference U minus V and proceed to add and subtract the product U minus V times M times U minus V-Hermitian in the right hand side of (12). Doing so and reordering terms, we rewrite E_i with the expression we show in (13). Some intermediate manipulations are needed to see that (13) is correct. You can do them on your own if you want to be thorough.

7. We now take norms in this expression. Using the triangle inequality and the sub multiplicative property of matrix operator norms, we obtain a version of (13) in which all of the factors are replaced by their norms. This is the expression we show in (14).

8. In (14), we have that the norm of M is the norm of E, which we know is epsilon. That the norms of these two matrices are the same is true because M is the eigenvalue matrix of E. We also have that the norm of V equals one. This is true because V is an orthonormal matrix, that satisfies V times V-Hermitian equals identity. We then have that (14) reduces to the expression shown in (15).

9. The factor between square brackets is an alternative form of the eigenvector misalignment constant delta that we defined in (4). You can verify that this is correct by expanding the square in Equation (4).

10. In any event. This is what we wanted to show.

11. Lemma 1 decomposes the product E times v i in two terms. One of them, the first term m_i times v_i is aligned with the eigenvector v_i. This represents a perturbation of the eigenvalue lambda_i of S. This is because the change is aligned with the corresponding eigenvector, v_i.

12. The second term is the part of the perturbation that is not aligned with v_i. We an therefore see that it represents a perturbation of the eigenvector v_i itself. Both of these perturbations are small because they are both of order epsilon. The eigenvector perturbation is further multiplied by delta giving the claim that the norm of E_i is less than epsilon delta. This is as stated in (9), which is the claim the Lemma makes.

From Shift Perturbations to Filter Perturbations

1. (Empty)

2. Starting with the proof proper, our first task is to translate a shift operator perturbation into a filter output perturbation. We do this in this section. There are no steps here that are conceptually challenging. It is just a matter of performing the right algebraic manipulations.

3. Our first action is to replace the operator distance modulo permutation between the two filters H of S hat and H of S with the regular operator, distance between operators P_0 transpose H of S hat P and H of S.

4. We can do this because the operator distance modulo permutation is a minimum over all permutation matrices. Consequently, it has to be smaller than the regular operator distance obtained by P sub zero. Which is a particular choice of permutation matrix.

5. We now recall that since the filter H of S hat its permutation equivariant, the permutation and the shift operator can exchange places, therefore we can further bound the operator distance modulo permutation as shown in (17), where in the rightmost term we have the regular operator distance between the filters, H of P_0 transpose S hat P_0 minus H of S.

6. Thus, to prove the theorem, it suffices to compare the filter H of P_0 transpose S hat P0 with the filter H of S.

7. We can process this further by noting that in the definition of the distance between S and S hat back in equation (2), we have written P_0 transpose S hat P 0 as S plus ES plus SE where E. Where E is the given relative error matrix. We can therefore, write the difference between the filter, H of P 0 transpose S hat P 0 and the filter H of S, as the difference between the filter H of S plus ES plus SE and the filter H of S.

8. We have therefore concluded that it surfaces to bound the norm of the filter difference H of S plus ES, plus SE minus H of S. There are no permutations left in (18) and we have the perturbation explicitly written in terms of the error matrix E. Proceeding to the use of the definition of graph filters as polynomials on the shift operator, or a series on the shift operator to be more precise, we can write this difference explicitly as the difference of two series. Both of them utilizing coefficients h_k. But one of them utilizing the shift operator S plus ES plus SE as a variable. And the other series utilizing the shift operator S as a variable.

9. To move forward with the proof we need to expand the matrix power S plus ES plus SE to the power of k. We are going to do so to first order on E. Namely, by considering only the terms that are linear on the error matrix E and grouping all other terms in a matrix O sub k of E. The expansion of this matrix power to first order takes the form we show in (20). This is a matrix binomial expansion that you can check on your own. In doing so, you have to pay attention to the fact that the matrices S and E do not need to commute with each other. The product E times S is not the same as the product S times E. This is the reason why we have summands where the error E appears between two different powers of S.

10. After checking that it is correct, we can substitute the expression in (20) into the expression in (19). The terms that involve S to the power of k cancel out and we are left with the expression we show in (21). Where we have the filter coefficients h_k scaling the sums that appear in (20).

11. In (21) the term O of E, which is the sum of the terms O sub k of E modulated by respective filter coefficients h_k, is of second order because the filter's frequency response is an analytic function. This means that the limit in (22) is finite. As the norm of the error goes to zero, the norm of the matrix O of E vanishes at a square rate. This makes it negligible with respect to the other terms in Equation (21) which, we will see, go to zero at a linear rate.

12. Comparing (22) and (21), with the theorem's claim in Equation (7), we conclude that the theorem can be proven if the first term in the right hand side of (21), the one that is highlighted in red, is bounded as in equation (7).

13. Indeed, define the filter perturbation Delta of S to represent this term.

14. And substitute back from (23) into (21), (18) and (17) to conclude that (24) holds. More slowly, substitute the definition of Delta of S in (23). Into the filter difference in (21). Which equals the filter difference in (18). Whose norm bounds the operator distance modulo permutation between the filters H of S hat and H of S as is shown in (17). To conclude that the filter distance we want to bound is bounded by the sum of the norn of Delta of S with the norm of O of E.

15. The term O of E we have already seen is of order epsilon squared in Equation (22). We will show in the rest of this proof that the operator norm of Delta of S is bounded by: Two. Times C. Times one plus delta square-root-of-n. Times epsilon. Where C is the Lipschitz constant of the filter. delta is the eigenvector misalignment constant. And n is the number of nodes of the graph. And Epsilon is the norm of the error matrix E. Once this is shown, the proof will be complete.

Shifting to the GFT Domain

1. The second section of the proof involves shifting the analysis to the GFT domain.

2. It is, indeed, the time for us to shift the analysis into the graph Fourier Transform Domain. Recall then that we are using v_i to denote the eigenvectors of S. From the definition of the inverse GFT. We know that we can write an input signal x as the product of the eigenvector matrix V and the GFT x-tilde. This product can be alternatively written as the product of x_i tilde times v sub i. Where the x_i-tilde are the components of the GFT of x and the v sub i are the eigenvectors of S as we have just recalled. What is most important to remember in this equation is that the coefficients x_i tilde are the components of the GFT of the input signal x.

3. Using this expression for x, we can write the product Delta of S times x as a summation of products of the form Delta of S times v sub i. Each of them modulated by the respective graph Fourier transform coefficient x_i-tilde. Further using the definition of Delta of S in (23) yields the expression in (25).

4. This expression looks complicated. But this is only because it has several terms. In its derivation, we have just replaced Delta of S for its definition in (23) and used the GFT of the signal x to write the product Delta of S times x as a summation of products of the form Delta of S times v_i. Each of them multiplied by the GFT coefficients x_i-tilde

5. Considering now that v sub i is an eigenvector of S we have that S to the power of k-minus-r times v_i equals lambda_i to the power of k-minus-r times v_i. And, likewise, that S to the k-minus-r-plus-1 times v_i equals lambda_i to the k-minus-r-plus-1 times v_i. By using these facts we can simplify (25) to the expression in (26) . Where, really, all we have done is replace powers of S by powers of lambda_i.

6. We are now going to use the decomposition we sated in the eigenvector perturbation lemma. We are going to do that because in (26) a term of the form E times v_i has appeared. And in the eigenvector perturbation lemma, we have decomposed this product of E and v_i as the summation of a product between the scalar m sub i with v_i and a product between a matrix E sub i with v_i. Utilizing this decomposition in (26), we end up with the terms in (27) and (28), which are the ones we need to study to conclude the proof of the theorem.

Fact 1

1. We are going to prove two facts, the first fact is related to the first term, which is relatively easy to handle.

2. This fact, which we will call Fact 1, says that if we let Delta_1 of S times x represent the term in (27), its norm can be bounded by: Two. Times C. Times epsilon. As we show in (29). This bound holds for any vector x that has unit norm.

Fact 2

1. The second term is a little more difficult to handle. We will prove that the following fact holds.

2. If we denote the term in Equation (28) as Delta_2 of S times x, its norm can be bounded by: Two. Times C. Times delta. Times the square-root-of-n. Times epsilon. As we show in (30). This bound holds for any vector x that has unit norm. And this fact we are going to call it Fact 2.

Putting Fact 1 and Fact 2 Together

1. Let us see how we can put Fact 1 and Fact 2 together.

2. The difficult part of the proof is to prove Fact 1 and Fact 2. Assuming they hold, the rest of the proof of the theorem is about putting pieces in place. We have already proven in (23) that the operator norm modulo permutation of the difference between the filters H of S hat and H of S, which is the quantity we want to bound, is upper bounded by the sum of the norms of Delta of S and O of E.

3. And the term O of E we have already seen is of order epsilon squared in Equation (22).

4. If we assume they hold, it is ready to use Fact 1 and Fact 2 to bound the operator norm of Delta of S. To that end, we use the definition of the operator norm to write the operator norm of Delta of S as the maximum over all vectors with unit norm of the regular norm of the Delta of S times x.

5. If we further use the expression for Delta of S times x that we derived in (27) and (28) along with the triangle inequality for the maximization operation, we can further bound the operator norm of Delta of S as two separate maximizations. Both of them over vectors of unit norm. But one of them maximizing the norm of the expression in (27). And the other one maximizing the norm of the expression in (28).

6. But these are also the expressions that appear in Fact 1 and Fact 2. They correspond to the definitions of Delta_1 of S times x and Delta_2 of S times x. We can therefore use Fact 1 to bound (33). And we can use Fact 2 to bound (34). The result is a term Two-C-epsilon coming from (33) bounded as in Fact 1. And a term Two-C-delta-square-root-of-n-epsilon coming from (34) bounded as in Fact 2.

7. To conclude the proof substitute the bound in (35) into the bound in (31). Recall again that the norm of O of E is of order epsilon squared, as stated in (21). The proof is complete because these substitutions yield the theorem’s claim in Equation (7).

Proof of Fact 1

1. We move on to prove Fact 1.

2. Proving Fact 1 and Fact 2 are the key steps in the proof of the stability theorem. This is where we are going to use the integral Lipschitz conditions to bound error terms. Of the two facts, Fact 1 is the easier to prove. We repeat the statement here for ease of reference.

3. The statement of Fact 1 involves the definition of the term Delta_1 of S times x and the bounding of its norm by Two-C-epsilon for any vector x that has unit norm. In the definition of Delta_1 of S times x, we have three summations. The innermost sum is over index r, which represents different powers of S. It comes from the binomial expansion in Equation (20). The middle sum is over index k, which is associated with different filter coefficients h_k. It comes from the definition of the graph filter we introduced in Equation (19). The outermost sum is over index i, which is associated with the components of GFT of x. It comes from writing the signal x with the inverse GFT. As we did in Equation (25).

4. To begin with the proof of Fact 1, we focus first on the innermost sum, the one over index r. And notice that the error matrix eigenvalue m sub i is a scalar that can change places with powers of S. We now use again the fact that v sub i is an eigenvector of S to write S to the power of r times v_i as lambda_i to the power of r times v_i. And, likewise, to write S to the power of r-plus-1 times v_i as the product of lambda_i to the power of r-plus-1 time v_i. When we do that, the innermost sum in (36) reduces to the expression we show in (37). Where, really, all we have done is replace powers of S by powers of lambda_i, which we can do because lambda_i is the eigenvalue of the matrix S associated with eigenvector v sub i.

5. This substitution produces a remarkable simplification. Because the sum in (37) involves a term of the form lambda_i to the k-minus-r times lambda to the r, which equals lambda_i to the k, irrespective of the value of r. And a second term of the form lambda_i to the k-minus-r-plus-1 times lambda_i to the r-plus-1 which reduces to lambda_i to the power of k irrespectively of r as well.

6. Thus, all of the terms in the sum are raised to the same power, k, irrespectively of the index r. We therefore have a total of k terms, each of which is equal to lambda_i, raised to the power of k plus lambda_i raised to the power of k. We can therefore reduce (37) to the simple expression in (38), where the sum is replaced by a factor of the form Two

times k times lambda_i to the power of k. This substitution is possible because, as we have explained, this is the value that the sum takes.

7. We can now substitute the expression in (38) back into the middle of the sum in (36), the one that is over indexes k associated with different filter coefficients h_k. This substitution is implemented in the first equality in (39). We then reorder terms to obtain the second equality in (39). We pull the factors 2 and m_i, which do not depend on the summation index k, as common factors

8. The manipulations we started in (37) and are about to complete are the key steps in the proof of Fact 1.

9. The reason why these steps are key is that we have ended with an expression that is: First of all, unexpectedly simple. And, second, suspiciously similar to the derivative of the frequency response of the filter. Indeed, remember that we can write the filter's frequency response as a series with coefficients h_k on powers of lambda. Therefore, the derivative satisfies the relationship we show in (40) where lambda_i times the derivative h-prime of lambda_i is written as the exact same series that appears in (40). This is, by the way, the same observation we made when studying dilations of the shift operator S. The two proofs are conceptually identical because of this. Except, of course, that reaching the point at which this observation becomes useful is much easier in the case of dilations.

10. Continuing with the proof, Equation (40) is substituted into equation (39). Where the series is replaced by the product lambda_i times h-prime of lambda_i. These two quantities being equal because of (40).

11. This ends the core part of the proof of Fact 1. The rest is just algebra. You should realize that this is true because the quantity we highlight in red in (41) is something we can bound with the integral Lipschitz condition. The bound appears in Equation (6), if you want to check it.

12. This is possible indeed. But it has to be checked. To that end, consider the norm squared of Delta_1 of S times x. It's energy, if you prefer. In Equation (42), the first equality comes from substituting (41) into (36). This eliminates the two innermost sums in (36) and leaves us with the outermost sum only. The one that is over GFT indexes i. In this sum, the eigenvectors v_i are orthogonal and have unit norm. Therefore, each term of the sum is the side of a **right** simplex. Then, the theorem of Pythagoras holds and we can write the squared norm of the sum as the sum of the squares of the individual summands.

13. Observe now that in (42), the eigenvalues of the error matrix are all of them bounded by epsilon. This is just the definition of a norm. The norm of a matrix is its largest eigenvalue. All of the eigenvalues m sub i of the error matrix E are smaller than its largest. Which is the norm epsilon.

14. Further note that the factors lambda_i times h prime of lambda_i are bounded by C. This comes from the integral lipschitz hypothesis on the frequency response of the filter. As stated in Equation (6).

15. We can therefore bound the operator norm of Delta_1 of S as we show in (43). Where m_i squared is bounded by epsilon squared and lambda_i times h prime of lambda_i squared is bounded by C squared.

16. To complete the proof of Fact 1 recall that the graph Fourier transform preserves energy. Therefore, the sum of the GFT components in (43), which is the energy of the GFT of the signal x equals the energy of x itself. Which is one because we are considering the computation of an operator norm.

17. Take square root on both sides of (43)

18. To conclude the proof of Fact 1.

19. Notice how working with energies here instead of norms is crucial in obtaining a bound that does not depend on the number of notes n. We could have used the triangle inequality in (42), but that would have prevented us from taking advantage of the orthogonality of the eigenvectors v sub i. We took advantage of that by working with square norms or energies and invoking Pythagoras's theorem.

20. This observation is not crucial to understand the proof of Fact 1. But it is an interesting observation nevertheless. The observation is also significant because the same cannot be done in the proof of Fact 2. The impossibility of invoking Pythagoras's theorem is the reason why the square-root-of-n term appears in Fact 2. Conversely, the possibility of invoking Pythagoras's theorem in this proof is the reason why a square-root-of-n term does not appear in Fact 1.

Proof of Fact 2

1. (Empty)

2. We move on to the proof of Fact 2. Which we repeat here for ease of reference.

3. Fact 2 involves the definition of the term Delta_2 of S times x and the bounding of its norm by Two times C times delta times the square-root-of-n times epsilon. This bound holds for any vector x that has unit norm. In the definition of Delta_2 of S we have, as we had in the definition of Delta_1 of S, three sums. One over index r representing powers of S. One over index k, representing filter coefficients h_k. And another one over index i, representing graph Fourier transform entries.

4. We begin the proof of Fact 2 with a focus on the innermost two sums. We write the eigenvector decomposition of S as V times lambda times V-Hermitian. And bring the eigenvector matrix V to the front of the sum. This allows us to write the expression in (45). Notice how inside of the sum, powers of S in the left hand side are replaced by powers of Lambda in the right hand side.

5. The term within brackets. The one that is nested between V and V-Hermitian, is a diagonal matrix. This is true because it is the sum of the diagonal matrices Lambda to the power of r and Lambda to the power of r-plus-1 multiplied by scalars. We define a matrix to represent this term. Since the matrix depends on the eigenvalue lambda_i, we will denote it as G sub i. We state this definition in Equation (46).

6. Using the definition of G_i the equality in (45) can be rewritten as we show in (47). This is just using the definition. The result looks misleadingly simple. The matrix G_i has a complicated expression.

7. The key step in the proof of Fact 2 is to manipulate the entries of G sub i to show that complicated thought they look, the entries of G_i actually have simple expressions. And that, more importantly, we can bound the entries with the integral Lipschitz condition. To work on these manipulations we write the diagonal entries of G sub i explicitly in Equation (48).

8. Recall that since G_i is a diagonal matrix these ae the only nonzero entries that this matrix has. To obtain (48) from (46) we replace the eigenvalue matrix Lambda with the eigenvalue lambda_j wherever that Lambda appears. We do this because we are looking at the j-th diagonal entry of G_i and lambda_j is the j-th diagonal entry of the eigenvalue matrix.

9. To continue with the processing of the entries of G_i we differentiate the cases in which j equals i and in which j is different from i.

10. For the case in which j equals i, in the expression for the entry G_i sub j-j in (48) we are actually looking at the entry G_i sub i-i as we show in (49). In this particular case the

terms inside the sum in Equation (48) are such that i and j are the same. Thus, the product lambda_i to the k-minus-r times lambda_j to the r is just lambda_i to the power or k. Likewise, the product lambda_i to the k-minus-r-plus-1 times lambda_j to the r-plus-1 reduces to lambda_i to the power of k, because indexes i and j are the same here.

11. Both of these facts hold true for all r. Thus, as was the case of the proof of Fact 1, compare the derivation we are doing here with Equation number (40), the derivative of the frequency response makes an appearance. We have that the i-th entry of the matrix G sub i is given by Two times lambda_i times h prime of lambda_i as we show in (49).

12. For the case in which j and i are different, the innermost sums in (48), the ones that are over index r, are geometrics sums. This allows for their computation in closed form. We show the resulting value of the sum in (49). You can compute these geometric sums on your own to verify that (49) is correct. Or not even. It's not very relevant to understand the proof.

13. What is relevant, is to use the explicit form in (49), back in (48). When we do that, we end up with sums that make the frequency response of the filter appear. They are the terms h of lambda_i, which comes from the summation of h_k times lambda_i to the power of k and h of lambda_j, which comes from the terms involving the sum of h_k times lambda_j to the power of k. Therefore we can reduce the j-th diagonal entry of the matrix G_i to the expression we show in (50).

14. It is time for us to make the crucial observation in the proof of Fact 2. That observation is that the integral Lipschitz hypothesis applies to both, the term that appears in (49) and the terms that appear in (51). The bound in Equation (6) applies to (49) and the bound in Equation (5) applies to (51). Therefore, the norm of the matrix G_i, which been diagonal is the absolute value of its largest entry, is bounded by Two times C, as we show in (52).

15. This completes the core of the proof of Fact 2, the rest is some simple algebra of matrix norms.

16. However, we do need to verify that this is true, so let us do that. We begin by taking norms in (47) and leveraging the sub multiplicative property of the operator norm. The norm of the product of matrices in (47) is thus rewritten as the product of the individual norms of each factor. This gives us the expression in (53).

17. In (53), the norm of the matrix V and the norm of the vector v_i are units. Because the matrix V is unitary and the eigenvector v_i is normalized. The norm of the matrix G_i is

bounded by 2C. This is the bound we have endeavored to establish. It is stated in (52). This is one of the critical observations to make in (53). The other critical observation to make in (53) is that the norm of E_i is bounded by the eigenvector perturbation lemma. That's the reason why we derived this lemma. The specific claim is that the norm of E_i is, at most, epsilon times delta. The bound is stated in Equation (9). We can therefore bound (53) as in (54)

18. At this point, we must recall that we have been concentrating in the two innermost sums that appear in the definition of Delta_2 of S in Equation (44). We need to add back the sum over GFT indexes. Making it so, our manipulations lead us to the bound we show in (55).

19. We now use the triangle inequality to bound the norm of the sum in (55) as the sum of norms of each summand. We end up with terms that have the form that appears in (54). Using the bound we show in this equation, allows us to write the bound in (56).

20. We now need to observe that the sum of the absolute values of the GFT components that appears in (56) is the 1-norm of the graph Fourier transform x tilde. The 1-norm of any vector of a given dimension is bounded by the 2-norm multiplied by the square root of the dimension. We can therefore bound the 1-norm of the GFT in (56) with its 2-norm multiplied by the square-root-of-n. This is where the square-root-of-n term appears. We then obtain the bound we show in (57).

21. Recall that the graph Fourier transform preserves energy and that the signal x under consideration has unit energy. Therefore, the 2-norm of the GFT coincides with the 2-norm of the signal x. Which is one because we are considering the computation of an operator norm.

22. This concludes the proof of Fact 2 and concludes the proof of theorem one.