What Makes Writing in Mathematics "Good"?

Intrinsic Dimensionality Estimation via Singular Value Curves

D.R. Hundley[*]
Mathematics Department
Whitman College
Walla Walla, WA 99362
hundledr@whitman.edu

M.J. Kirby[†]
Mathematics Department
Colorado State University
Fort Collins, CO 80521
kirby@math.colostate.edu

Abstract

We propose a modification and a generalization of a method for determining the intrinsic dimensionality (ID) of data, the method of singular value (SV) curves. While other methods use a single computation of the covariance of a ball of data, this method takes multiple measurements of a ball of increasing radius, and is therefore itself a generalization of the widely used PCA spectrum. Our modification is geared towards making the SV curves canonical, allowing for an automation of the ID estimation. We then generalize the technique to give an analytic version of the SV curves, making them suitable to analyze noisy data sets. We give several examples and show that the extra computational effort is well worthwhile when one desires an in-depth analysis of the local topological structure of the data.

1 Introduction

Local representations of smooth data are linear. In fact, they result directly from the well-known Taylor's formula in m-dimensions which provides a basis for the tangent space about the point of approximation. This tangent space is the column space of the Jacobian of the surface function which, in data analysis problems, is generally unknown. Here we want to be explicit about our assumptions on the data- We will assume that a given (local) data set, has the following decomposition,

$$\boldsymbol{x} \in \mathbb{R}^n \Rightarrow \hat{\boldsymbol{x}} \in \mathbb{R}^k$$

where there exists a (perhaps nonlinear) function f so that $\boldsymbol{x} = f(\hat{\boldsymbol{x}})$ and k is the smallest such integer. Thus, our $\hat{\boldsymbol{x}}$ is a $k-$dimensional parametrization of the surface given as data points $\boldsymbol{x}$, and k is the intrinsic dimension of the data. Of course, in practice, the equality given above is only an approximation.

We note that this is in contrast to other definitions of intrinsic dimension, notably from [1], where the definition is given as the number of dimensions necessary to *encapsulate* the data- if the intrinsic dimension of the data is k by this definition, then the data lies entirely

[*] http://www.whitman.edu/˜hundledr
[†] http://www.math.colostate.edu/˜kirby

What “Good Writing” Usually Does

- Definitions before use.

What “Good Writing” Usually Does

- Definitions before use.
- *Why*.

What "Good Writing" Usually Does

- Definitions before use.
- *Why*.
- Say why, then how (structure)

What “Good Writing” Usually Does

- Definitions before use.
- *Why*.
- Say why, then how (structure)
- Sentences to connect equations and justify steps.

What “Good Writing” Usually Does

- Definitions before use.
- *Why*.
- Say why, then how (structure)
- Sentences to connect equations and justify steps.
- Notation consistent and readable.

Example.

Let G be a graph and D a dominating set. We find bounds and characterize the extremal cases. This is useful and interesting and has many applications.

Example.
Let G be a graph and D a dominating set. We find bounds and characterize the extremal cases. This is useful and interesting and has many applications.

Problems

Example.
Let G be a graph and D a dominating set. We find bounds and characterize the extremal cases. This is useful and interesting and has many applications.

Problems

- "Dominating set" is used but never defined.

Example.
Let G be a graph and D a dominating set. We find bounds and characterize the extremal cases. This is useful and interesting and has many applications.

Problems

- "Dominating set" is used but never defined.
- "Find bounds" and "characterize" are too vague for a reader.

Example.

Let G be a graph and D a dominating set. We find bounds and characterize the extremal cases. This is useful and interesting and has many applications.

Problems

- "Dominating set" is used but never defined.
- "Find bounds" and "characterize" are too vague for a reader.
- "Useful and interesting" does not tell the reader *why* or *where*.

Example re-write

Example.
Let G be a finite graph. A *dominating set* for G is a subset D of vertices such that every vertex of G is either in D or adjacent to a vertex in D.

Dominating sets arise naturally in applications such as network monitoring, where one wishes to place the smallest number of sensors so that every node is observed.

In this paper, we investigate bounds on the size of minimal dominating sets for trees and provide a constructive characterization of extremal cases.

Example re-write

Example.
Let G be a finite graph. A *dominating set* for G is a subset D of vertices such that every vertex of G is either in D or adjacent to a vertex in D.

Dominating sets arise naturally in applications such as network monitoring, where one wishes to place the smallest number of sensors so that every node is observed.

In this paper, we investigate bounds on the size of minimal dominating sets for trees and provide a constructive characterization of extremal cases.

What works well?

Example re-write

Example.
Let G be a finite graph. A *dominating set* for G is a subset D of vertices such that every vertex of G is either in D or adjacent to a vertex in D.

Dominating sets arise naturally in applications such as network monitoring, where one wishes to place the smallest number of sensors so that every node is observed.

In this paper, we investigate bounds on the size of minimal dominating sets for trees and provide a constructive characterization of extremal cases.

What works well?

- Term is defined before use.

Example re-write

Example.
Let G be a finite graph. A *dominating set* for G is a subset D of vertices such that every vertex of G is either in D or adjacent to a vertex in D.

Dominating sets arise naturally in applications such as network monitoring, where one wishes to place the smallest number of sensors so that every node is observed.

In this paper, we investigate bounds on the size of minimal dominating sets for trees and provide a constructive characterization of extremal cases.

What works well?

- Term is defined before use.
- Reader gets an application-level reason to care.

Example re-write

Example.
Let G be a finite graph. A *dominating set* for G is a subset D of vertices such that every vertex of G is either in D or adjacent to a vertex in D.

Dominating sets arise naturally in applications such as network monitoring, where one wishes to place the smallest number of sensors so that every node is observed.

In this paper, we investigate bounds on the size of minimal dominating sets for trees and provide a constructive characterization of extremal cases.

What works well?

- Term is defined before use.
- Reader gets an application-level reason to care.
- Clear statement of paper scope and direction.

Example re-write

Example.
Let G be a finite graph. A *dominating set* for G is a subset D of vertices such that every vertex of G is either in D or adjacent to a vertex in D.

Dominating sets arise naturally in applications such as network monitoring, where one wishes to place the smallest number of sensors so that every node is observed.

In this paper, we investigate bounds on the size of minimal dominating sets for trees and provide a constructive characterization of extremal cases.

What works well?

- Term is defined before use.
- Reader gets an application-level reason to care.
- Clear statement of paper scope and direction.

Bad Example 2: Abrupt / Compressed Proof

Proposition. If n is even, then n^2 is divisible by 4.

Proof. n is even so $n = 2k$. Then $n^2 = 4k^2$ which is divisible by 4. Done.

Bad Example 2: Abrupt / Compressed Proof

Proposition. If n is even, then n^2 is divisible by 4.

Proof. n is even so $n = 2k$. Then $n^2 = 4k^2$ which is divisible by 4. Done.

Problems

Bad Example 2: Abrupt / Compressed Proof

Proposition. If n is even, then n^2 is divisible by 4.

Proof. n is even so $n = 2k$. Then $n^2 = 4k^2$ which is divisible by 4. Done.

Problems

- Too compressed: gives the reader no guidance.

Bad Example 2: Abrupt / Compressed Proof

Proposition. If n is even, then n^2 is divisible by 4.

Proof. n is even so $n = 2k$. Then $n^2 = 4k^2$ which is divisible by 4. Done.

Problems

- Too compressed: gives the reader no guidance.
- "Done" is not a mathematical conclusion.

Bad Example 2: Abrupt / Compressed Proof

Proposition. If n is even, then n^2 is divisible by 4.

Proof. n is even so $n = 2k$. Then $n^2 = 4k^2$ which is divisible by 4. Done.

Problems

- Too compressed: gives the reader no guidance.
- "Done" is not a mathematical conclusion.
- Missing connective statements: why does each step follow?

Bad Example 2: Abrupt / Compressed Proof

Proposition. If n is even, then n^2 is divisible by 4.

Proof. n is even so $n = 2k$. Then $n^2 = 4k^2$ which is divisible by 4. Done.

Problems

- Too compressed: gives the reader no guidance.
- "Done" is not a mathematical conclusion.
- Missing connective statements: why does each step follow?

Bad Example 3: Equation Dumping

Example.

$$a_{n+1} = \frac{1}{2}a_n + 1, \quad x = \frac{1}{2}x + 1 \Rightarrow x = 2, \quad a_{n+1} - 2 = \frac{1}{2}(a_n - 2), \quad |a_n - 2| = 2^{-n}|a_0 - 2| \to 0.$$

Bad Example 3: Equation Dumping

Example.

$$a_{n+1} = \frac{1}{2}a_n + 1, \quad x = \frac{1}{2}x + 1 \Rightarrow x = 2, \quad a_{n+1} - 2 = \frac{1}{2}(a_n - 2), \quad |a_n - 2| = 2^{-n}|a_0 - 2| \to 0.$$

Problems

Bad Example 3: Equation Dumping

Example.

$$a_{n+1} = \frac{1}{2}a_n + 1, \quad x = \frac{1}{2}x + 1 \Rightarrow x = 2, \quad a_{n+1} - 2 = \frac{1}{2}(a_n - 2), \quad |a_n - 2| = 2^{-n}|a_0 - 2| \to 0.$$

Problems

- Minimal prose: reader must guess purpose of each line.

Bad Example 3: Equation Dumping

Example.

$$a_{n+1} = \tfrac{1}{2}a_n + 1, \quad x = \tfrac{1}{2}x + 1 \Rightarrow x = 2, \quad a_{n+1} - 2 = \tfrac{1}{2}(a_n - 2), \quad |a_n - 2| = 2^{-n}|a_0 - 2| \to 0.$$

Problems

- Minimal prose: reader must guess purpose of each line.
- No interpretation of the equations.

Bad Example 3: Equation Dumping

Example.

$$a_{n+1} = \frac{1}{2}a_n + 1, \quad x = \frac{1}{2}x + 1 \Rightarrow x = 2, \quad a_{n+1} - 2 = \frac{1}{2}(a_n - 2), \quad |a_n - 2| = 2^{-n}|a_0 - 2| \to 0.$$

Problems

- Minimal prose: reader must guess purpose of each line.
- No interpretation of the equations.
- Misses the chance to highlight the key idea (contraction toward a fixed point).

Bad Example 3: Equation Dumping

Example.

$$a_{n+1} = \frac{1}{2}a_n + 1, \quad x = \frac{1}{2}x + 1 \Rightarrow x = 2, \quad a_{n+1} - 2 = \frac{1}{2}(a_n - 2), \quad |a_n - 2| = 2^{-n}|a_0 - 2| \to 0.$$

Problems

- Minimal prose: reader must guess purpose of each line.
- No interpretation of the equations.
- Misses the chance to highlight the key idea (contraction toward a fixed point).

Quick Checklist for Student Drafts

When revising, ask:

1. Did I define every nonstandard term and symbol *before* using it?
2. Did I tell the reader what the goal is (theorem, classification, method, example set)?
3. Does each proof have a visible structure (strategy + justified steps)?
4. Do my equations have "glue sentences" that explain what they mean?
5. Could a peer summarize each paragraph's purpose in one sentence?

Wrap Up

- In pairs, rewrite the "bad" examples into "good" examples.
- Review each other's work
- Rewrite.
- Turn in your papers at the end of class.