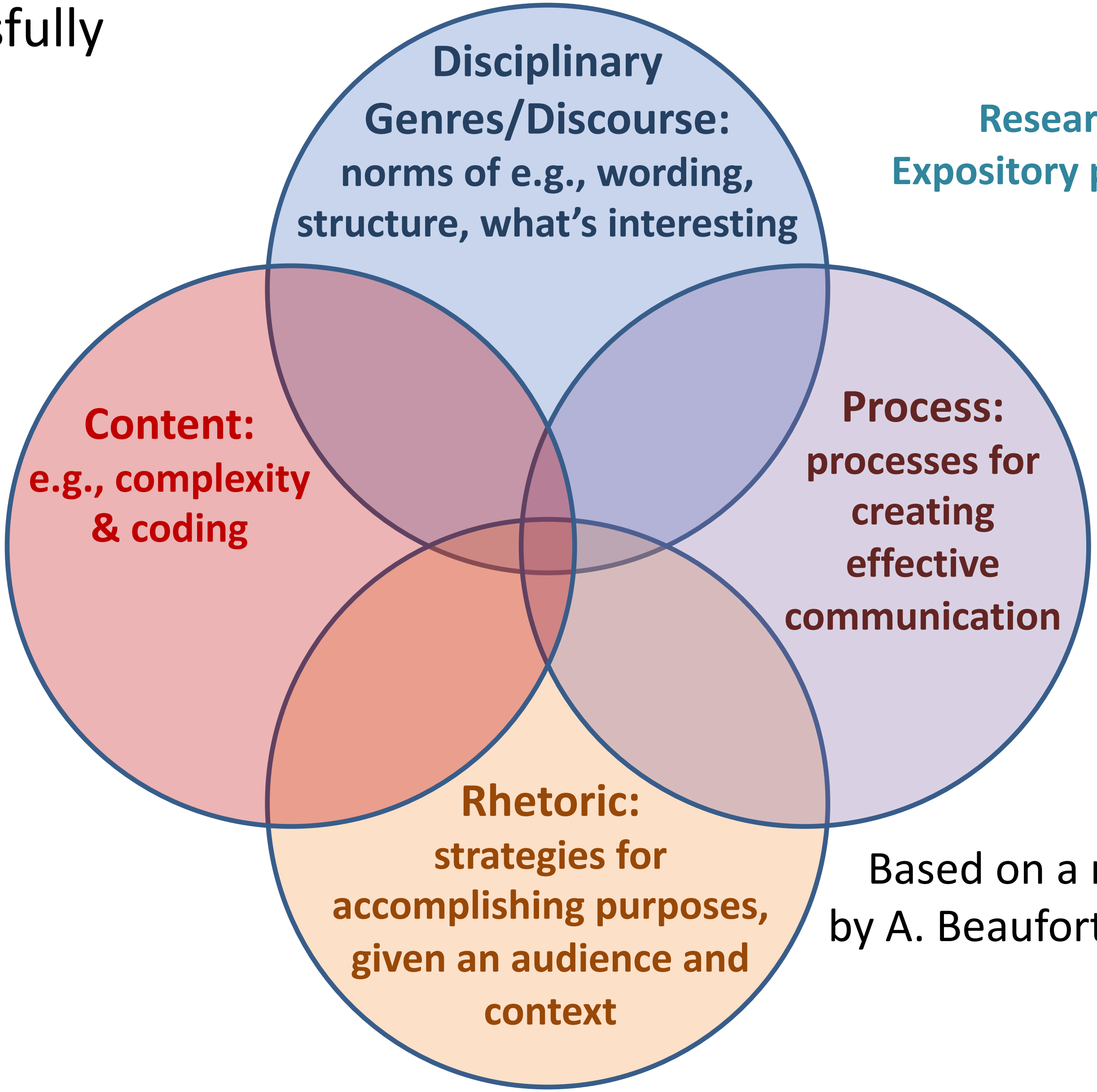
Crafting text to help readers follow the logic

18.424 Spring 2023

Susan Ruff
ruff@math.mit.edu

Communicating successfully in a new discipline requires expertise in several domains of knowledge



Lecture note: teaches
Research paper: presents new results
Expository paper: presents existing results (e.g. 18.424 term paper)

Based on a model by A. Beaufort 2007

Today: Communicating clearly to peers in a research paper or expository paper.

GUIDING TEXT

- Ensure readers know
  - WHAT you're doing
  - WHY you're doing it
  - HOW you're doing it

- ~~**Always tell** readers~~
  - ~~WHAT you're doing~~
  - ~~WHY you're doing it~~
  - ~~HOW you're doing it~~

III. PROOF OF THEOREM 2 FOR A SPECIAL CASE

In this section we prove Theorem 2 for the very special case discussed in Section I. All alphabets $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$ are equal to $\{0, 1\}$. The source $\{S_k\}$ satisfies $\Pr\{S_k = 0\} = \Pr\{S_k = 1\} = \frac{1}{2}$. Channel Q_M is noiseless, i.e., $Q_M(y|x) = \delta_{x,y}$; and channel Q_W is a BSC with crossover probability p_0 $(0 \leq p_0 \leq \frac{1}{2})$, i.e.,

$$Q_W(z|y) = (1 - p_0)\delta_{y,z} + p_0(1 - \delta_{y,z}). \tag{17}$$

We show here that (R, d) is achievable if and only if

$$R \leq C_M = 1, \quad d \leq H_S = 1, \quad Rd \leq h(p_0). \tag{18}$$

Since, for this case, $\Gamma(R) = h(p_0)$, this result is a special case of the as-yet-unproven Theorem 2. We begin with the converse ("only if") part of the result. Let $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Z}^N$ correspond to an encoder-decoder (N, K, Δ, P_e) (note that $\mathbf{Y}^N = \mathbf{X}^N$). Then, making repeated use of the identity $H(U, V) = H(U) + H(V|U)$, we can write (dropping the superscript on vectors)

$$\begin{aligned} K\Delta &= H(\mathbf{S}^K|\mathbf{Z}^N) = H(\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &= H(\mathbf{S}, \mathbf{X}, \mathbf{Z}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &= H(\mathbf{Z}|\mathbf{X}, \mathbf{S}) + H(\mathbf{X}, \mathbf{S}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &\overset{(a)}{=} H(\mathbf{Z}|\mathbf{X}) + H(\mathbf{S}|\mathbf{X}) + H(\mathbf{X}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &\overset{(b)}{=} Nh(p_0) + H(\mathbf{S}|\mathbf{X}) + [H(\mathbf{X}) - H(\mathbf{Z})] - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}). \end{aligned} \tag{19}$$

These steps are justified as follows.

The Wire Tap Channel, by A.D. Wyner

Crafting “flow” Gopen & Swan 2018, 1990

Topic Verb Point of Closure

I ate cookies yesterday.

Known Action **New**

Crafting "flow"

Topic — Point of Closure

I ate cookies yesterday.

Known — New

Readers expect **Known** ⟶ **New** information order.

Using Known→New information order has a nice side effect:

A→B. B→C. C→D.

Tight connectivity between sentences creates "flow," which can make the text easier to read and follow.

To describe a broom, we recall the wave packet decomposition of Ef introduced by Bourgain [1]. The wave packet decomposition says that inside a large ball of radius R, we can decompose Ef into a sum over wave packets $Ef_{\theta,v}$. Each wave packet $Ef_{\theta,v}$ is essentially supported in a tube $T_{\theta,v}$ of length R and radius $R^{1/2+\delta}$ for some small $\delta > 0$. The axis of $T_{\theta,v}$ points in a direction depending only on θ, and the location of $T_{\theta,v}$ is described by v. The absolute value $|Ef_{\theta,v}|$ of a wave packet is approximately a constant function on $T_{\theta,v}$.

A Restriction Estimate in R3 using Brooms, by Hong Wang

Crafting “flow”

Topic — Point of Closure

I ate cookies yesterday.

Known — New

Readers expect

Known ⟶ New

information order.

Using Known→New information order has a nice side effect:

A→B. B→C. C→D.

Tight connectivity between sentences creates “flow,” which can make the text easier to read and follow.

But for writers, it’s natural to do otherwise:

A→B. C→D.

Crafting "flow"

Topic Point of Closure

I ate cookies yesterday.

Known **New**

Readers expect

Known ⟶ **New**

information order.

Using Known→New information order has a nice side effect:

A→B. B→C. C→D.

Tight connectivity between sentences creates "flow," which can make the text easier to read and follow.

But for writers, it's natural to do otherwise:

A→B. C→D.

A break in flow can be caused by

missing information

Some conceptual gaps are ok, so long as the reader doesn't experience them as a gap.

1. Plugging our initial values into (2) we see the expression simplifies to 2+2. Four is less than…

2a. Plugging our initial values into (2) we see the expression simplifies to 2+2. Perfect squares have the property…

2b. Plugging our initial values into (2) we see the expression simplifies to 2+2, which is a perfect square. Perfect squares have the property…

Activity by Malcah Effron

Crafting “flow”

Topic — Point of Closure

I ate cookies yesterday.

Known — **New**

Readers expect

Known ⟶ **New**

information order.

Using Known→New information order has a nice side effect:

A→B. B→C. C→D.

Tight connectivity between sentences creates “flow,” which can make the text easier to read and follow.

But for writers, it’s natural to do otherwise:

A→B. C→D.

A→B. C←B.

A break in flow can be caused by

missing information

poor information order

You try it.

The Catalan numbers C_n are known to count the number of Dyck paths of length $2n$, that is $C_n = |\mathcal{D}_n|$. ____________________________. Therefore, full binary trees are also enumerated by the Catalan sequence.

Fact: There exists a bijection between the set of full binary trees with k internal vertices and the set of Dyck paths of length $2k$.

Activity by Malcah Effron

You try it.

A common question in combinatorics is how to count, or enumerate, objects. For instance, how many possible full binary trees are there with N internal vertices? Since Dyck paths are enumerated by the Catalan numbers, the number of full binary trees with N internal vertices is C_N.

Fact: There exists a bijection between the set of full binary trees with k internal vertices and the set of Dyck paths of length $2k$.

Activity by Malcah Effron

Try it as you revise your short paper:

Topic   Point of Closure

I ate cookies yesterday.

Known   **New**

Readers expect

Known ⟶ **New**

information order.

Using Known→New information order has a nice side effect:

A→B. B→C. C→D.

Tight connectivity between sentences creates "flow," which can make the text easier to read and follow.

But for writers, it's natural to do otherwise:

A break in flow can be caused by

A→B. C→D.

missing information

A→B. C←B.

poor information order

Look for breaks in flow as you *revise*.

Tight connections between sentences create flow

A→B. B→C. C→D.

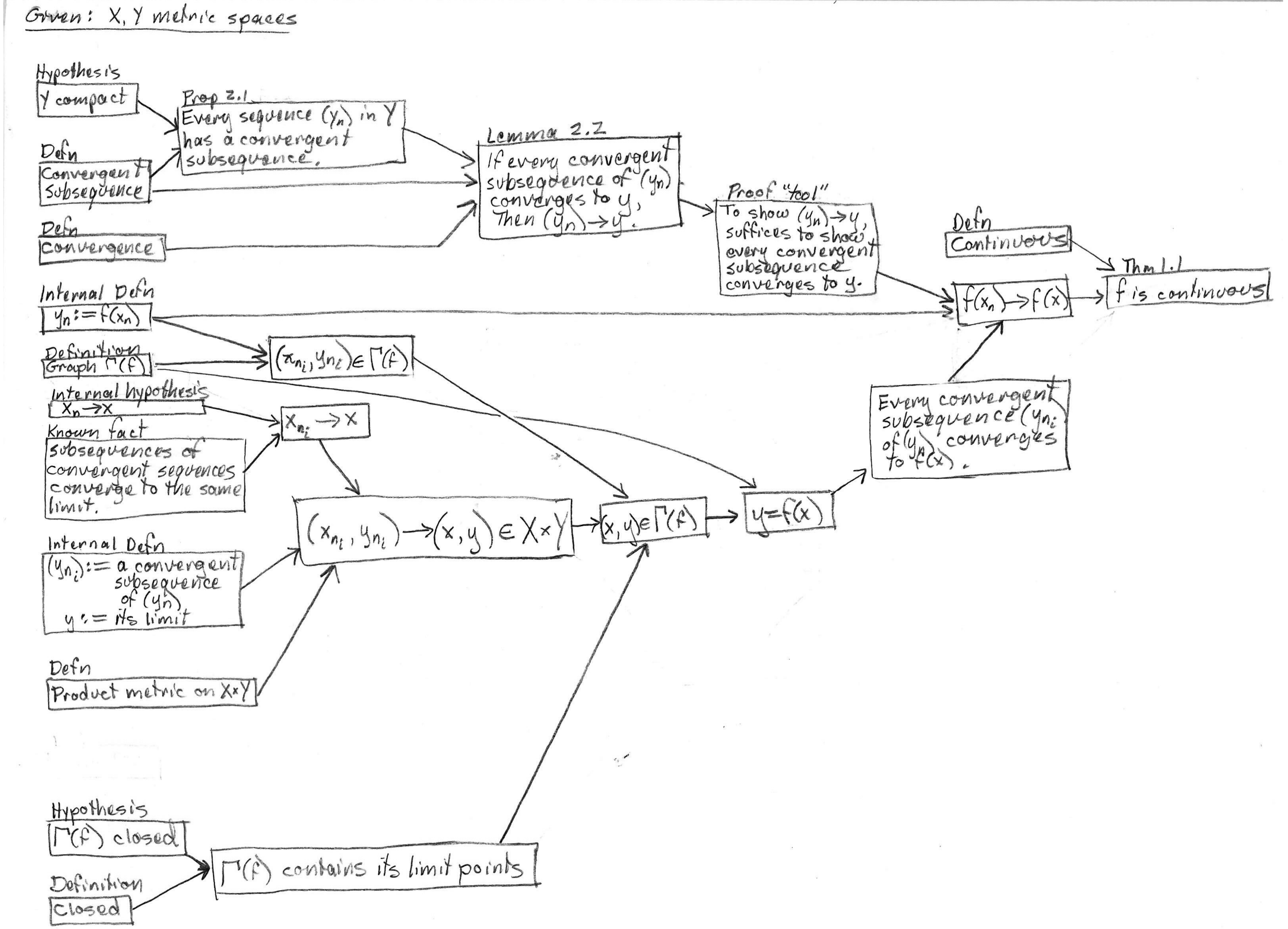
Here, **Blue = known**; **Red = new**

To describe a broom, we recall the wave packet decomposition of Ef introduced by Bourgain [1]. The wave packet decomposition says that inside a large ball of radius R, we can decompose Ef into a sum over wave packets $Ef_{\theta,v}$. Each wave packet $Ef_{\theta,v}$ is essentially supported in a tube $T_{\theta,v}$ of length R and radius $R^{1/2+\delta}$ for some small $\delta > 0$. The axis of $T_{\theta,v}$ points in a direction depending only on θ, and the location of $T_{\theta,v}$ is described by v. The absolute value $|Ef_{\theta,v}|$ of a wave packet is approximately a constant function on $T_{\theta,v}$.

A Restriction Estimate in R3 using Brooms, by Hong Wang

Text is inherently one dimensional: one long thread of text. But logic is not!

A proof's logic



One thread could use known-to-new flow.

How can we pull the various threads together?

Connect threads via Guiding text

- Ensure readers know
  - WHAT you're doing
  - WHY you're doing it
  - HOW you're doing it

III. PROOF OF THEOREM 2 FOR A SPECIAL CASE

In this section we prove Theorem 2 for the very special case discussed in Section I. All alphabets $\mathcal{S}$, $\mathfrak{X}$, $\mathcal{Y}$, $\mathfrak{z}$ are equal to $\{0, 1\}$. The source $\{S_k\}$ satisfies $\Pr\{S_k = 0\} = \Pr\{S_k = 1\} = \frac{1}{2}$. Channel Q_M is noiseless, i.e., $Q_M(y|x) = \delta_{x,y}$; and channel Q_W is a BSC with crossover probability p_0 $(0 \leqq p_0 \leqq \frac{1}{2})$, i.e.,

$$Q_W(z|y) = (1 - p_0)\delta_{y,z} + p_0(1 - \delta_{y,z}). \tag{17}$$

We show here that (R, d) is achievable if and only if

$$R \leqq C_M = 1, \quad d \leqq H_S = 1, \quad Rd \leqq h(p_0). \tag{18}$$

Since, for this case, $\Gamma(R) = h(p_0)$, this result is a special case of the as-yet-unproven Theorem 2. We begin with the converse ("only if") part of the result. Let $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Z}^N$ correspond to an encoder-decoder (N, K, Δ, P_e) (note that $\mathbf{Y}^N = \mathbf{X}^N$). Then, making repeated use of the identity $H(U, V) = H(U) + H(V|U)$, we can write (dropping the superscript on vectors)

$$\begin{aligned} K\Delta &= H(\mathbf{S}^K|\mathbf{Z}^N) = H(\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &= H(\mathbf{S}, \mathbf{X}, \mathbf{Z}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &= H(\mathbf{Z}|\mathbf{X}, \mathbf{S}) + H(\mathbf{X}, \mathbf{S}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &\overset{(a)}{=} H(\mathbf{Z}|\mathbf{X}) + H(\mathbf{S}|\mathbf{X}) + H(\mathbf{X}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\ &\overset{(b)}{=} Nh(p_0) + H(\mathbf{S}|\mathbf{X}) + [H(\mathbf{X}) - H(\mathbf{Z})] - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}). \end{aligned} \tag{19}$$

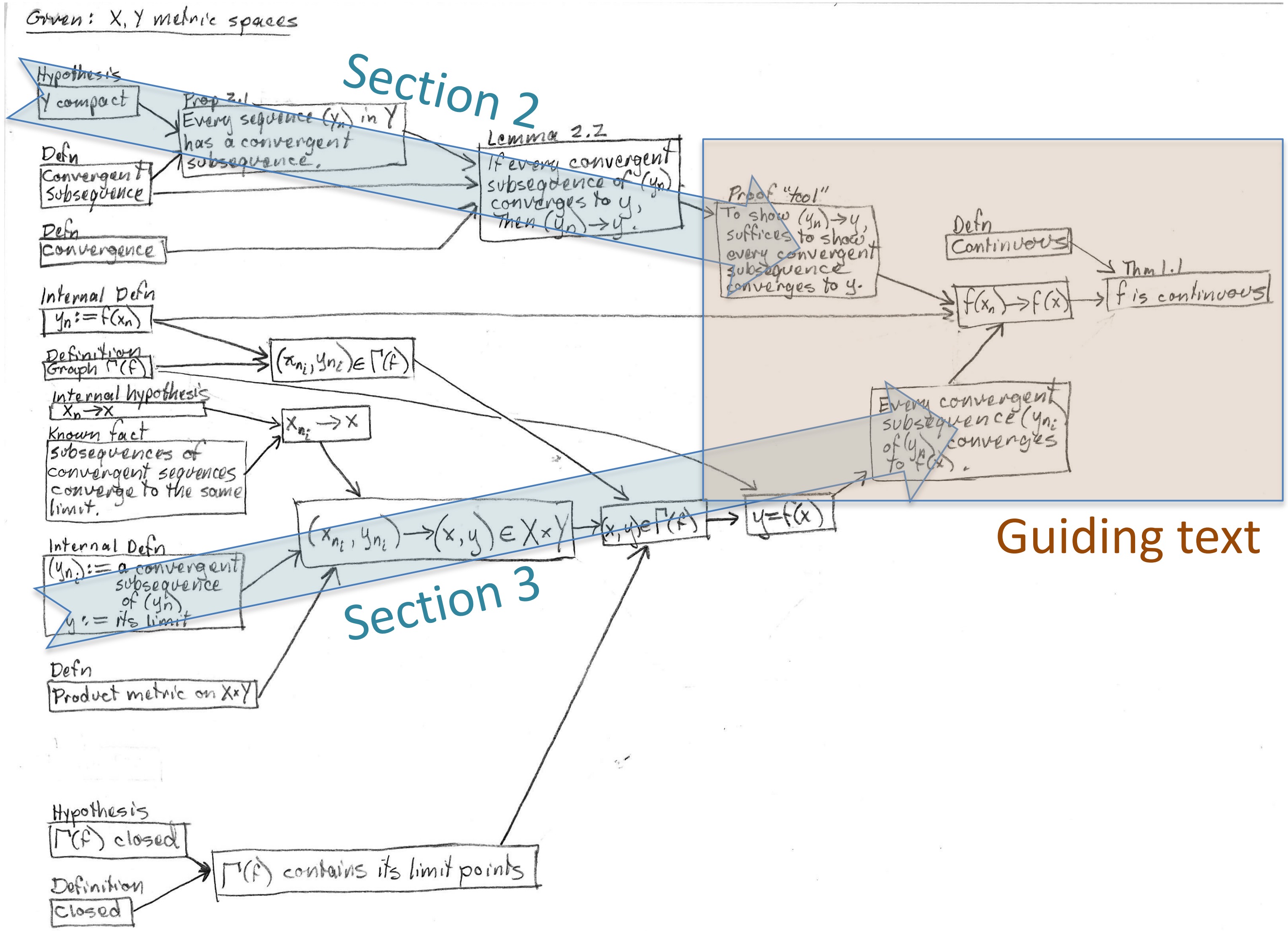
These steps are justified as follows.

The Wire Tap Channel, by A.D. Wyner

Connect threads via Guiding text

Guiding text:
To show f is continuous,
we prove in Section 2 that
it suffices to show…;
then in Section 3...

Guiding text enables
known-to-new
at larger scales



Also connect threads via key terms

a result of the polynomial partitioning, the function Ef can be split into a cellular term, a transverse term, and a tangential term (see Section 2.4). The cellular and the transversal contributions are estimated similarly, using the induction hypothesis on mass and radius. The tangential term is estimated directly, without appealing to induction.

The unconditional estimate for the tangential term remains favorable if the induction hypothesis is changed to accommodate Theorem 1.5, which is reflected in reducing the exponent weight on the L^2-mass. However, in this new setup, the estimate for the cellular part is no longer an immediate consequence of induction on radius. Most of the novelty of this paper goes into finding a new way to deal with the cellular contribution. Our argument contains a multistep iteration. And the scale of the

A Restriction Estimate on R^3 using Brooms, by Hong Wang

Exact repetition of key terms creates connection across distance.

As the introduction summarizes the paper's "story" it can set up key terms to later be "known"

large). Our main problem is the characterization of the family of achievable (R, d) pairs, and such a characterization is given in Theorem 2. It turns out (Theorem 3) that, in nearly every case, there exists a "secrecy capacity," $C_s > 0$, such that (C_s, H_S) is achievable [while, for $R > C_s$, (R, H_S) is not achievable]. Thus, it is possible to reliably transmit information at the positive rate C_s in essentially perfect secrecy.

An outline of the remainder of this paper now follows. In Section II, we give a formal statement of the problem and state the main results (Theorems 2 and 3). In Section III we give a proof of Theorem 2 for the special case discussed above (main channel noiseless, wire-tap channel a BSC). In Section IV, we prove the converse half of Theorem 2, and in Section V the direct half of that theorem.

The Wire Tap Channel, by A.D. Wyner

Guiding text and key terms work at all levels of discourse: paper level, section level, paragraph level, etc.

III. PROOF OF THEOREM 2 FOR A SPECIAL CASE

In this section we prove Theorem 2 for the very special case discussed in Section I. All alphabets $\mathcal{S}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathfrak{z}$ are equal to $\{0, 1\}$. The source $\{S_k\}$ satisfies $\Pr\{S_k = 0\} = \Pr\{S_k = 1\} = \frac{1}{2}$. Channel Q_M is noiseless, i.e., $Q_M(y|x) = \delta_{x,y}$; and channel Q_W is a BSC with crossover probability p_0 $(0 \leqq p_0 \leqq \frac{1}{2})$, i.e.,

$$Q_W(z|y) = (1 - p_0)\delta_{y,z} + p_0(1 - \delta_{y,z}). \tag{17}$$

We show here that (R, d) is achievable if and only if

$$R \leqq C_M = 1, \quad d \leqq H_S = 1, \quad Rd \leqq h(p_0). \tag{18}$$

Since, for this case, $\Gamma(R) = h(p_0)$, this result is a special case of the as-yet-unproven Theorem 2. We begin with the converse ("only if") part of the result. Let $\mathbf{S}^K$, $\mathbf{X}^N$, $\mathbf{Z}^N$ correspond to an encoder-decoder (N, K, Δ, P_e) (note that $\mathbf{Y}^N = \mathbf{X}^N$). Then, making repeated use of the identity $H(U, V) = H(U) + H(V|U)$, we can write (dropping the superscript on vectors)

$$\begin{aligned}
K\Delta &= H(\mathbf{S}^K|\mathbf{Z}^N) = H(\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\
&= H(\mathbf{S}, \mathbf{X}, \mathbf{Z}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\
&= H(\mathbf{Z}|\mathbf{X}, \mathbf{S}) + H(\mathbf{X}, \mathbf{S}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\
&\overset{(a)}{=} H(\mathbf{Z}|\mathbf{X}) + H(\mathbf{S}|\mathbf{X}) + H(\mathbf{X}) - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}) - H(\mathbf{Z}) \\
&\overset{(b)}{=} Nh(p_0) + H(\mathbf{S}|\mathbf{X}) + [H(\mathbf{X}) - H(\mathbf{Z})] - H(\mathbf{X}|\mathbf{S}, \mathbf{Z}).
\end{aligned} \tag{19}$$

These steps are justified as follows.

The Wire Tap Channel, by A.D. Wyner

Beware “superficial flow”

The market-determined price of a bond with fixed, known cash flows determines the bond’s internal rate of return, or yield. Different yields are typically approximately equal. Approximations can be provided by Taylor series. The Taylor series is due to James Gregory of Scotland. Scotland has 790 islands, including the Northern Isles and the Hebrides, according to Wikipedia. Wikipedia occasionally asks for donations.

“FLOW” SHOULD HELP READERS FOLLOW THE FLOW OF THE LOGIC

Takeaway: Craft your paper so

- Known-to-new structure
- Guiding text
- Key terms

all work together to create a cohesive paper

that reveals the flow and structure of the underlying logic.