Mathematical Methods

Li-Chung Wang

December 10, 2011

Abstract

In this paper, we try to classify all the mathematical methods. Mathematical methods are too many to count. They look too diversified to manage. If we attempt to classify them by luck, it would be difficult as though we were lost in a maze and tried to get out. However, if we caredully study a method's origins, development and functions, we may find some clues about our task by following along their texture. Studying methods helps us analyze patterns. In order to make a method's essence outstanding, the setting must be simple. Using a method is like producing a product: A method can be divided into three stages: input, process, and output. Suppose we compare Method A with Method B. If the input of Method B is *more than or equal to* that of Method A and if both the process and the output of Method B are better than those of Method A, then we say that Method B is more *productive* than Method A. Here are some examples. Cases when the input is increased: The method given in the general case can be applied to specific cases. However, specific cases contain more resources, there may be more effective methods available for specific cases. Our goal is to seek the most effective method in each case. Strategies for improving the output: When we fomulate a theorem, the conclusion should be as strong as possible. When we seek a solution, the solution should be specific and precise. The execution of the solving plan should be finished perfectly and the solution should be expressed in closed form if possible. Strategies for improving the process: When we introduce a definition, the definition should be accessible in a finite number of steps; the target should be reached quickly, directly and effectively. When we present a proof or theory, we should avoid repetition, orgnize it in a consistent and systematic way, make it leaner and simpler. The qualities of process can be roughly divided into following categories: Accessibility, functions (reduction, analogy, effectiveness, simplicity, directness, flexibility, fully utilizing resources, hitting multiple targets with one shot, avoiding contradictions, avoiding repetition and unnecessary complications, expanding the scope of application without loss of efficiency), structurization (insightfulness, essence, flowcharts or networks). Structurization may help us systematically operate and deeply understand the complicated cases. Like two neighboring colors in a rainbow, there is no clear dividing line between two qualities. The method of *weakening a hypothesis* is a method of *reducing* the input for a given output. The concept of almost everywhere can often ruin the delicacy of the method of weakening a hypothesis. For physical methods, we discuss physical interpretations and physical proofs. As we go to a more advanced level and widen our consideration, new physical meanings of mathematical equations continue to develop and meanings of equations become richer and more delicate. Nonetheless the meanings in older theories are still well-preserved in a newer theory. Guided by a theorem's physical meaning, one may develop a better strategy to prove it. The ideal physical proof is the one each of whose step has a pertinent physical interpretation. The development of physical methods shows the tendency toward such an ideal. For example, the mathematical formulation of the second law of thermodynamics leads to a criterion for integrability of Pfaffian forms. A physical proof is usually more direct than a geometric proof. Physical and geometric proofs provide richer meanings, insights, and interesting stories than analytic proofs. A proper physics model can be a natural guide to the study of PDEs. Improvements of classical methods: a remedy rather than a thorough revamp is all we need; this introduction mode based on needs may make the key to

improvement most outstanding. A hyperbolic paraboloid has center at (∞,∞,∞) because Cartesian coordinates lack ability to distinguish infinities of all directions; if we use spherical coordinates instead, there will be no common midpoint for the chords through the origin. If the set of centers is empty and we allow a point involving ∞ to be its element due to a tool abuse, then all the theorems to which the false existence of elements leads will be meaningless. Only after a construction is tailored to our needs can it solve the problem effectively. An index set must be chosen properly: one more candidate would be too many and one less would be too few. The Dirac delta generalized function should not be treated as a function; we provide an easy way to bridge the gap between a function and a generalized function. A good theorem should provide complete information. The Ritz method is an effective tool for studying Sturm–Liouville Problems. In order to effectively solve a problem, we must quickly understand the circumstance with the minimum effort, and then directly attack the heart of the matter. Find extrema with subsidiary conditions. Compare physics proofs with mathematics proofs. Distribution theory is a new theory that we create to avoid the contradiction that the domain of a function contains a point whose function value cannot be defined. How do we deal with a problem that may easily cause us to commit errors?

When studying a generalized definition, we should understand its primitive version, its entire process of revolution, and the reason for the necessity of generalization; if we proceed directly toward the most general version in axiomatic approaches, its setting usually requires a more strange language and less familiar structures which may blur the essential idea, and the algorithm to check the definition usually becomes less effective; thus, an improper approach to generalized definitions may easily lead to an empty formality and make it difficult for us to see the advantages of generalized definitions over the primitive version; providing several non-trivial examples alone is not enough.

"Using formulas in a table without care may easily result in mistakes. One is under the impression that once the solution form is obtained, the actual solution is determined. This is not so. If the resulting function is multivalued and the formula fails to indicate which value to choose, then the formula would be useless. One should find a delicate method to determine the correct value. If one uses such a unfinished formula in a proof, then the proof would be incorrect. Such a mistake is often difficult to detect."

"How do we detect errors in a textbook? When I find an error, the first response is usually to refuse to accept this fact and try to rationalize the opposite viewpoint. After all, there are many authors who have not found it incorrect after copying it. Nevertheness, I try to remember this odd experience so that I can easily find a reason when a problem occurs afterwards. However, this "rationalization" actually conceals a mistake. The reason why we fail to detect an error is that we have not gone far enough to forsee its consequences. Errors cannot withstand tests. Soon or later they will be detected. Even if an error may not be detected at the first checkpoint in application, it can hardly survive at the second one. If we consider an error true, then the world would fall into pieces as if Pandora's box were opened. I became so frustrated that I had to choose the other option: the error is not necessarily true. Then I found a counterexample. I could omit some details and still make this paragraph logical, but this would destroy the evidence of true experience and eliminate the track of the natural thought for solving a problem."

"Contour integrals for special functions: I. When we deal with a contour integral for a special function, all we have to do is to choose a point on the contour and assign a possible value to its argument. II. The only purpose of detailed discussion about branch points is to tell us that if we want to choose a point and its argument properly to facilitate calculations, we must consider branch points first."

"Tying up loose ends"

"The finishing touch: Providing a solution to a problem alone is not enough; the author should tell the readers from where the solution comes. This way can bring the readers to an advantageous point for a bird's-eye view of the circumstance."

"Musket to kill a butterfly: Example [Differentiating under an integral]. The modern method attacks directly toward the goal by using therems flexibly. A complex measure need not distinguish a compact integral contour from a noncompact one. A single proof is good enough for dealing with both compact and noncompact cases. Furthermore, the proof is free from complex analysis except for using the definition of analytic functions. In contrast, the classical method must follow a formal, tedious, and inflexible procedure. In order to ensure the finiteness of a contour integral, the Borel measure must distinguish a compact integral contour from a noncompact one. In fact, in order to include the case of noncompact integral contour, the modification and supplement have to use almost all the theorems in complex analysis and, thus, lead to unnecessary complications."

"Grasping the overall situation: Hypergeometric functions and confluent hypergeometric functions are closely related. We must build paths between the two topics as many as possible. When we discuss confluent hypergeometric functions, of course, we have to include their characteristic properties. Furthermore, for each property, we should find its corresponding property in hypergeometric functions, treat the latter as a motivation of the former and use the latter to prove the former. Just because of the complicated circumstance, we should give a rigorous proof rather than touch it lightly. Otherwise, the discussion is incomplete.Sneddon [74, p.32, l.1–l.18] sets a good example for discussing confluent hypergeometric functions."

"Linear transformations of the hypergeometric function: the general equation of Fuchsian type having three regular singularities vs. the standard hypergeometric equation; eliminating repetitions in an overestimate vs. the counting based on the correspondence between solution pairs and regular singularities; Guo–Wang [39, p.141, (4) & (5)] can be derived from Guo–Wang [39, p.140, (2) & (3)] by inspection [Watson–Whittaker [88, p.207, (I) & (II)]]; calculations vs. inspection; Lebedev [52, §9.5] shows that Lebedev [52, p.249, (9.5.8) & (9.5.9); p.250, (9.5.10)] all follow from Lebedev [52, p.249, (9.5.7); p.247, (9.5.1) & (9.5.2)]; based on the list of linear transformations given in Lebedev [52, p.246, l.-14], the discussion given in Lebedev [52, §9.5] is complete; the formula given in Watson–Whittaker [88, p.289, l.3–l.5] and the one given in Watson–Whittaker [88, p.291, l.3–l.5] are proved the hard way because they both use the contour integrals of Barnes' type [p.286, l.-7–p.287, l.3; p.289, l.-18–l.-17] and the residue theorem; in fact, we can still prove Lebedev [52, p.247, (9.5.1) & (9.5.2); p.248, (9.5.4); pp.249–250, (9.5.7)–(9.5.10)] without using any integral representation."

"Methodical solutions: First, consider a differential equation of a special type. If its integral solution is based on guess, luck, and trial-and-error, we do not know from where the integrand comes, and the only way to justify the solution is by substitution, then this underdeveloped solution cannot be consider a methodical solution. Suppose the same equation also belongs to the wider class of equations of Laplacian type. In contrast, its integral solution can be built by a systematic method. In fact, the integrand and the path of integration can be specified by the Laplace transform. Consequently, the latter solution is more methodical than the former one."

"Applications of analytic continuation to the Weber–Schafheitlin integral (the right timing for a statement's appearance): Suppose we choose the weakest possible conditions required in an argument to be our theorem's hypothesis. If the argument has used the method of analytic continuation no more than once, then no confusion will occur. However, what should we do if the argument has used the method of analytic continuation more than once?

Proposing a new condition without collecting enough evidence in advance has a problem with the timing for its appearance. Therefore, whenever we use the method of analytic contiuation, we should check and record if the change of the condition is needed so that we may easily clarify the relationship between cause and effect in the proof structure."

"Integration on a Riemann surface with branch points: If we reduce a contour integral on a Riemann surface to an integral along a line segment, the value of the latter integral may depend on which sheet the line segment is in, while the former integral is an invariant quantity. When we reduce a contour integral on a Riemann surface to an integral along a line segment, we often have to degeneate a part of the contour to a point. In order to make the argument of points along the contour continuous and simplify the calculation of these arguments, we should restore the degeneated point to its corresponding nondegeneate part."

"Contour integrals for Bessel functions"

"The recurrance formulas for Neumann's polynomials given by Watson [89, p.274, (1), (2) & (3)] can be derived from

I. The relation between Bessel coefficients and Neumann's polynomials: Watson [89, p.271, (1)],

II. The Laurent series expansion for Bessel coefficients: Watson [89, p.14, (1)], and

III. The recurrence formula for the generating function of Bessel coefficients: Watson [89, p.45, (1)] (see Watson [89, p.275, l.9–l.10]).

Remark 1. Want to prove uniform convergence when convergence is given

Remark 2. Series rearrangement
Remark 3. Detailed analysis."
"Determine $\arg(1-t)$ on a contour around the branch point $t=1$: We need a method rather than correct results. Any step coming from guess may lead to the desired result this time; it may not next time. For example, if the choice $\arg(-1)=\pi$ can lead to the desired result, we want to know why we cannot choose $\arg(-1)=-\pi$. Thus, if one provides correct results without a method, one may still make mistakes sometimes. Ten correct examples are not as good as one correct method. Only when a complete method is provided may we check if results are correct. When encountering a situation where a confusion may easily occur, we should deliberately clarify the confusion rather avoid discussing it."
"Binomial series: The classical view emphasizes the choice of principal value and the consistency with previous results. The modern view emphasizes whether $\sum_{k=0}^{\infty}\binom{\alpha}{k}z^k$ is convergent and whether the cases considered are inclusive."
"Listing examples cannot be considered a proof: Listing examples cannot be considered a proof just like a tangled ball of yarn cannot be called a piece of cloth. A professional proof must give the direction of thoughts and the key idea. We should not avoid discussing the part difficult to discribe. On the contrary, we should work harder to give it a clear explanation.
A finite series must have the first term, the last term, and the general term. An infinite series must have the first term and the general term. To figure out the general term of a series from its first few terms is an example of inductive reasoning or a conjecture, but should not be considered a proof. For a binomial coefficient, we should use its compact symbol $\binom{n}{k}$ rather than its awkard factorial form $\frac{n!}{k!(n-k)!}$ unless for the purpose of computation. The formulas given in Hobson [41, p.106, l.-8–l.-1] look messy due to the abuse of notation. Suppose n is even. Hobson [41, p.107, (7)] expresses $\cos n\theta$ as a finite series in ascending powers of sine without the highest power term. Hobson [41, p.105, (3)] expresses $(-1)^{n/2}\cos n\theta$ as a finite series in descending powers of sine without the lowest power term. Hobson [41, p.107, l.12] claims that Hobson [41, p.107, (7)] is Hobson [41, p.105, (3)], written in reverse order. How is it possible to compare two things when one of them is unknown?
Hobson [41, §78 & §79] expresses $\cos n\theta$ and $\sin n\theta$ as descending power series of sine. Their combinatorial proofs are tedious and annoying. If we want to express them in ascending power series of sine, all have to do is list all the terms of the descending power series and then reverse the order. However, Hobson [41, §80–§83] fails to do this simple way by repeating the same kind of tedious and annoying combinatorial proofs. Mathematics is not for killing time. We have more important things to do.
Clarification of a point of confusion"
"The proof of a theorem is hidden in the application whose proof requires the use of the theorem: Example. Stirling's theorem"
"Finding the inverse function of a given analytic function with the Fourier series method"
"Statements of a certain type have the same proof pattern"
"The Taylor series vs. the L'hopital rule in terms of convergence: Sometimes, only after studying advanced mathematics may we understand how we should properly deal with elementary mathematics. In order to study infinite products of analytic functions, we must master the concept of uniform convergence. Thus, it is important to see how the Taylor series and the L'hopital rule affect convergence. Among proofs for the case of point convergence, we should select the ones applicable to the case of uniform convergence."
"Differentiation of a rational function whose denominator is a high power of a polynomial: One may use the product rule $(fg)'=f'g+fg')$ or use the quotient rule $(\frac{f}{g})=\frac{gf'-fg'}{g^2}$. If one uses the latter rule, one should not expand the two terms in the numerator of the resulting rational function. Cancel the common factor of the numerator and the denominator of the resulting rational function first. This may avoid a lot of unnecessary computations. If one were to expand the terms in the numerator after using the quotient rule, this would make it difficult to cancel the common factor or identify the complicated resulting expression with the desired value."
"The motive of creation and process of evolution for the method of steepest descents"
"With vs without guess and check: Proofs are used to check the truth of a statement and are not neces-

sarily helpful to understand its meaning. For example, we can use the mathematical induction to prove $\sum_{k=1}^{n} k^2 = n(n+1)(2n+1)/6$, but do not know how we get this formula. The proof is independent of the theme of this formula just like a quality control inspector checks only the packaging of product. This is a proof with guess and check; its analysis for the formula is shallow. We make the conclusion without enough confidence beforehand, and have to check afterwards; the guess and poor explanation lowers the quality of theory. Therefore, ideal and mature mathematical theories should gradually eliminate the guesswork in it. The features of a proof without guess and check: having a specific viewpoint; starting with a careful plan to get the answer; all the operations being in control beforehand."

"Infinite integrals: Tests of convergence: the comparison test, Abel's test, and Dirichlet's test. Tests of uniform convergence: the method of change of variable, Abel's test, Dirichlet's test, and Lebesgue's dominated convergence theorem. Many theorems about uniform convergence can be considered corollaries of Lebesgue's dominated convergence theorem. We often evaluate infinite integrals by using Taylor series expansions.

The process of evolution for Abel's test for uniform convergence vs that for Weierstrass' test"

"A science book author should not use definitions to stop readers' questions: For any science book, a reader should not accept a definition as a command about whose origin one should not question although it does not require a proof. An author should not give a definition without providing a reason."

"The right timing for correcting mistakes: In physics, we study facts. Theories are nothing but tools to explain facts. When a theory fails to explain facts, it should be abandoned and eliminated. When we find a statement contradictory to facts, we should trace to the origin of mistake and rewrite the theory from there. Of course, an incorrect statement will lead to a lot of junks, but we are not interested in why they are junks. The important thing is to correct mistakes as soon as they occur. Perhaps the Gibbs paradox is valuable for books about the development history of stastistical mechanics, but not for a textbook. A textbook should not contain any incorrect statement because it is a reference book for quotation and application."

"Maxwell made a contradiction compatible by changing $\nabla \times H = J_f$ to $\nabla \times H = J_f + J_d$: the contradiction to be resolved, his analysis, his remedy for compatibility, how the correction of the formula affects the results whose validity depends on the formula, and other evidence of the existence of displacement current."

"Faraday made a contradiction compatible by changing $\nabla \times E = 0$ [static] to $\nabla \times E = -\frac{\partial B}{\partial t}$ [nonstatic]."

"How we should properly treat Ampèpere's law: the situation, our strategy, and the value of Ampèpere's law."

"A more delicate and effective method provides more information."

Keywords. Productive methods, accessibility, functions, reduction, analogy, effectiveness, simplicity, directness, flexibility, fully utilizing resources, hitting multiple targets with one shot, structurization, insightfulness, essence, networks, compatibility, unifications, avoiding repetition and unnecessary complications, method of weakening hypothesis, physical methods, physical interpretations, physical proofs, Leibniz integral rule, Cauchy's theorem, Cauchy's integral formula, residue theorem, Runge's theorem, Dirac delta function, Heaviside function, Green functions, Bessel functions, Riemann zeta function, Lipschitz conditions, boundary conditions, regular singular points, ratio test, root test, Frobenius method, inverse function theorem, Riemann–Lebesgue lemma, prime number theorem, Baire's category theorem, Pragmen–Lindelöf theorem, Paley–Wiener theorem, functional analysis, splitting field, perfect field, Lagrange's resolvents, Galois resolvents, recurrence relations, generating function, Lengendre's equation, integral transforms, Laplace transform, separation of variables, tensor product, wedge product, contravariant vectors, covariant vectors, fundamental groups, covering space, topological group homomorphism, covering group, evolute, involutes, centers of curvature, envelope, normals, normal incidence, singularity of the second kind, boundedness, growth condition, rotation operators, limit-point case, real nondecreasing spectral functions, Hermitian nondecreasing spectral matrices, Sturm's oscillation theorem, Prüfer substitution, Poincarè phase plane, varia-

tional derivative, isoperimetric problems, holonomic problems, non-holonomic problems, positive semidefinite matrix, canonical Euler equations, characteristic system of the Hamilton–Jacobi equation, Hamilton–Jacobi theory, method of characteristics, Galilean transformations, Lorentz transformations, Michelson–Morley experiment, Maxwell's equations, muon decay, special relativity, proper time, relativistic kinetic energy, relativistic Lagrangian, center-of-mass coordinate system, coupled harmonic oscillators, normal modes, generalized coordinates, exact differential, separation of variables, ruled surface, developable surface, directrix, argumented matrix, polar lines, polar planes, osculating plane, osculating circle, osculating sphere, the second law of thermodynamics, Pfaffian forms, random variable, characteristic function, distribution, inversion formula, convergence in probability, almost sure convergence, strong law of large numbers, central limit theorem, cluster point, Ritz method, Sturm–Liouville problems, direct methods, method of finite differences, method of Lagrange multipliers, equation of the vibrating membrane, vector triple product, testing functions, contour integrals, dummy variable, branch points, Euler transforms, Fuchsian type, Borel measures, hypergeometric, confluent, regular singularity, Riemann's P-equation, Lebesgue dominated convergence theorem, methodical solutions, Weber–Schafheitlin integral, Riemann surfaces, Neumann's polynomials, Bessel coefficients, Zhu–Vandermonde's identity, method of steepest descents, Weierstrass' test, Abel's test, Dirichlet's test, Cauchy data, Cauchy–Kowalevski theorem, microstates, entropy, extensive property, Gibbs paradox, displacement current, induced current, homopolar generator, magnetostatics, idealized circuit

Providing a proof without a method involves giving the final answer by intuition first and then justifying it from hindsight (see the proof of Munkres [57, p.322, Theorem 1.2]). We may distill a method from the above justification. Understanding the method will enable us to systematically proceed toward the solution by analyzing patterns and taking advantage of the circumstances. Munkres [58, p.327, l.7–l.10; l.11–l.14; p.328, l.$-$14–p.329, l.6] are parts of the method. A method is useful for generalization (Munkres [58, p.329, Theorem 51.3]).

A method is the summary of essential ideas for solving a problem. The solution guided by a method is often concise, organized, and insightful (Edwards [26, p.4, l.$-$2–p.5, l.13]).

1 The approach from the microscopic viewpoint vs. that from the macroscopic viewpoint

From the microscopic point of view, solving a problem is equivalent to exploring possibilities. From the macroscopic point of view, solving a problem is equivalent to eliminating impossibilities. For example, when we try to factor cyclotomic integers into ideal prime divisors, we may

(1) Use the approach given in Edwards [25, p.128, l.$-$3–p.129, l.21] to construct ideal prime divisors (explore the possibilities) or

(2) Use Stewart–Tall [78, p.186, Theorem 10.1] to find the ideal prime factorization in the general case and then prove that the $e_1, \cdots, e_r$ given in Stewart–Tall [78, p.186, l.8] are all equal to 1 using van der Waerden [82, vol.1, p.120, l.4–l.19] (eliminating the impossibilities).

2 How we choose the most suitable setting for illustrating a method

(Partitions of unity Munkres [57, p.222, Theorem 5.1])

The construction of a partition of unity has wide applications: topology, real analysis (Rudin [72, p.41, Theorem 2.13]), and differential geometry (Spivak [77, vol. 1, p.69, Corollary 16]). In essence, the construction of a partition of unity is a topological method. In order to ensure the method's wide application, the setting should be general. Dugundji [24, p.144, Proposition 3.2] and the diagram given in Dugundji [24, p.311] show that a locally compact, paracompact, or normal topological space meets the setting requirement. In order to expressively illustrate a method's essence, the construction should be simple. The method given in Munkres [57, p.222, Theorem 5.1] is simpler than that given in Rudin [72, p.41, Theorem 2.13]. The formulation and proof of Urysohn's lemma given in Dugundji [24, p.146, Theorem 4.1] is simpler than those given in Rudin [72, p.40, Proposition 2.12] and Spivak [77, vol. 1, p.44, Lemma 2]. Furthermore, choosing a finite partition of unity will free us from considering the nuisance given in Dugundji [24, p.170, Definition 4.1(1)]. Except for settings, the constructing methods in Munkres [57, p.222, Theorem 5.1], Dugundji [24, p.170, Theorem 4.2] and Spivak [77, vol. 1, p.68, Theorem 15] are the same. All the above considerations make normal spaces the best choice of a setting for illustrating "partitions of unity".

If we discuss theorems or solutions of differential equations, the method can be divided into three stages: input, process, and output. For a theorem, the input is the hypothesis and the output is the conclusion. For solving a differential equation, the input is the problem and the output is the solution. The output is an essential tool for determining the quality of a method in this case. If we discuss theories, proofs, or definitions, it sufficies to consider input and process because we are interested only in their method. The output stage can be ignored in this case. The input is the settings and the process is the formulations.

Suppose we compare Method A with Method B. If the input of Method B is more than or equal to that of Method A and if both the process and the output of Method B are better than those of Method A, then we say that Method B is more productive than Method A. The method given in the general case can be applied to specific cases. However, specific cases contain more resources, there may be more effective methods available for specific cases. Our goal is to seek the most effective method in each case.

3 Productive methods

3.1 Strategies for improving the output

When we fomulate a theorem, the conclusion should be as strong as possible. Specialization often divides the discussion of a topic into cases and reduces results to the simplest form for each case. When we seek a solution, the solution should be specific and precise. The solving plan should be executed thoroughly and perfectly; the solution should be expressed in closed form if possible.

Example 3.1.

Although the form given in Watson–Whittaker [88, p.365, l.2–l.3] is good for generalization (Watson–Whittaker [88, p.368, l-9–l.-1]), it is not as effective as the forms given in Guo–Wang [39, p.351, (5) & (6)]. First, the former form has not been reduced to simple form for each case, so it is not good for direct application. Second, if a series terminates, we want to know how many terms it has, otherwise the answer is not complete.

Example 3.2. (The construction of Green's functions in one dimension)

The definition of a Green's function is given by Ince [43, p.254, l.16–l.21]. The uniqueness of Green's function follows from Ince [43, p.254, l.19–l.20] (Coddington–Levinson [18, p.192, l.23–l.27]). Indeed, as a function of x, $G-\tilde{G}$ is of class C^{n-1} because G and $\tilde{G}$ have the same discontinuity at $x=\xi$. Although Green's function is unique, there are many methods for its construction. Here are some examples: Ince [43, p.254, l.-19–p.255, l.-11], Coddington–Levinson [18, p.190, l.28–p.192, l.1] (let $l=0$), Gerlach [33, Theorem 45.1]. The first example provides a solution which proposes a plan but fails to execute it for the treatment of both the differential equation and the boundary conditions. The second example provides a solution which finishes the treatment of the differential equation, but only proposes a plan for dealing with the boundary conditions without finishing the plan. The third example provides a solution which finishes the treatment of both the differential equation and the boundary conditions. The more precise the form is, the stroger properties of Green's function we may obtain from it (Compare Ince [43, p.254, l.16–l.21] with Coddington–Levinson [18, p.192, l.12–l.20]). The property given in Coddington–Levinson [18, p.192, l.14–l.15] can be used to prove the formula given in Ince [43, p.257, l.2] (Rudin [72, p.27, Theorem 1.34]). The reason given in Ince [43, p.257, l.1] is incorrect. Now we discuss the above three examples in detail.

(1) Corrections for the first example:

  (a) "$P(G,H)=0$ when $x=a$ and when $x=b$" given in Ince [43, p.256, l.7–l.8] should be replaced with "$P(G,H)|_a^b=0$ (Ince [43, p.213, l.19])".

  (b) $p_0[H\frac{d^{n-1}G}{dx^{n-1}}-G\frac{d^{n-1}H}{dx^{n-1}}]$ given in Ince [43, p.256, l.12] should be replaced with $p_0[H\frac{d^{n-1}G}{dx^{n-1}}+(-1)^{n-1}G\frac{d^{n-1}H}{dx^{n-1}}]$.

  (c) $p_0(\xi_1)H(\xi_1,\xi_2)\lim[\frac{d^{n-1}G}{dx^{n-1}}]_{\xi_1-\varepsilon}^{\xi_1+\varepsilon}-p0(\xi_2)G(\xi_2,\xi_1)\lim[\frac{d^{n-1}H}{dx^{n-1}}]_{\xi_2-\varepsilon}^{\xi_2+\varepsilon}=0$ should be relplaced with $p_0(\xi_1)H(\xi_1,\xi_2)\lim[\frac{d^{n-1}G}{dx^{n-1}}]_{\xi_1-\varepsilon}^{\xi_1+\varepsilon}+(-1)^{n-1}p_0(\xi_2)G(\xi_2,\xi_1)\lim[\frac{d^{n-1}H}{dx^{n-1}}]_{\xi_2-\varepsilon}^{\xi_2+\varepsilon}=0.$

  (d) "$p_0(\xi_1)\lim[\frac{d^{n-1}G}{dx^{n-1}}]_{\xi_1-\varepsilon}^{\xi_1+\varepsilon}=p_0(\xi_2)\lim[\frac{d^{n-1}H}{dx^{n-1}}]_{\xi_2-\varepsilon}^{\xi_2+\varepsilon}=1$" should be replaced with "$p_0(\xi_1)\lim[\frac{d^{n-1}G}{dx^{n-1}}]_{\xi_1-\varepsilon}^{\xi_1+\varepsilon}=(-1)^n p_0(\xi_2)\lim[\frac{d^{n-1}H}{dx^{n-1}}]_{\xi_2-\varepsilon}^{\xi_2+\varepsilon}=1$".

(2) Supplements of the second example: The proof of $Lu=lu+f$ can be found in Ince [43, p.256, l.-6–p.257, l.10]. The equality given in Ince [43, p.257, l.5] follows from the Leibniz integral rule and $\int_a^b=\int_a^{x^-}+\int_{x+}^b$.

(3) Supplements of the third example: Bernd [8] provides the motivation of the construction of Green's function given in Gerlach [33, Theorem 45.1]. Bernd [8, Example 3] relates Green's function to the Dirac delta function and the Heaviside function. Bernd [8, (5.35)] motivates us to generalize the relationships to the abstract level of functional analysis (Rudin [71, p.206, Exercise 10; p.378, l.-6]). Compare Bernd [8, (5.27)] with the formula given in Rudin [71, p.206, l.9].

3.2 Strategies for improving the qualities of process

The qualities of process can be roughly divided into following categories:

(1) Accessibility: Construct the existence of solution in a finite number of steps. Avoid using any proposition whose validity cannot be verified in a finite number of steps. Avoid using the axiom of choice, reduction to absurdity, and mathematical induction. If we must use reduction to absurdity or mathematical induction, we should narrow its scope of application wherever possible. When using mathematical

induction, we should reduce the amount of work in the induction step wherever possible lest the program takes takes too much time and memory in computer.
(Discussion) Some mathematicians think that existence can be established by construction, by the method of reduction to absurdity, or by assumption. Once the existence is established, we should not worry about the method of establishing the existence. However, mathematicians in the intuitionist school insist that only when every claim during the construction of existence can be determined to be true in a finite number of steps may the existence be considered mathematically significant.

A compound sentence is true only if each of its component sentences is true. If one of its component sentence cannot be determined to be true or false in a finite number of steps, then this compound sentence is mathematically meaningless. As for the choice of resources and tools, in principle, we choose only necessary ones (Edwards [26, p.68, l.21–l.26]). Discard irrelevant and unnecessary ones (Edwards [26, p.68, l.$-$14–l.$-$9]).

(2) Functions

(a) Reduction: Reduce calculations; reduce to lower-level systems. The canonical Euler equations represent the characteristic system associated with the Hamilton–Jacobi equation [Fomin–Gelfand [30, p.90, l.$-$12–l.$-$10]]. Originally, this fact was a part of Hamilton–Jacobi theory in classical mechanics. Since then the method of *characteristics* has been developed to be an important tool in *reducing* a *partial* differential equation to a system of *ordinary* differential equations. Read Tkachev [81, Method of characteristic strips] and Gibbon [34, chaps. 1 & 2].

(b) Analogy: We should link a new concept with a familiar one so that we have a model in mind for studying the new concept.
(Discussion) When we study a new concept, the first thing we should do is relate it to a familiar concept by establishing a major link between them. This is because analogy provides a vantage point to see the big picture. Before the link is established, every task is difficult. Once the link is established, every task becomes easy.

(c) Effectiveness: This quality refers to accessibility, specification, elementary methods, quantitative instead of qualitative formulations, the construction of solutions by an effective algorithm rather than trial and error, the use of simple theorems rather than complicated ones.
(Discussion) Although effective mathematics can use available resources to provide an effective method of constructing the strong existence of a mathematical object, it sometimes has congenital defects in other mathematical tasks. In general, if we try to emphasize efficiency, accuracy, concrete construction, an argument's strength, utilizing resources, or other details, we move toward effectiveness. However, if we try to emphasize the whole, we move away from effectiveness. Such tasks are unification, classification, abstraction, generalization, clarifying structures (Wang [85, Example 5.12 (The Jordan canonical form)]), identifying the essential reason for uniqueness, or proving the result is independent of our choices of construction.

(d) Hit multiple targets with one shot

(e) Directness: Adopt a direct approach to the solution rather than a roundabout one.

(f) Simplicity: Reduce the general form to a simple form. If there are several methods available, we should choose the simplest one. Sometimes one method is always simpler than others; sometimes the choice for simplicity varies from case to case.

(g) Avoid unnecessary complications: Avoid repetition or awkard languages. We should not unnecessarily generalize a theorem unless the generalized theorem has practical applications. Our argument should use simple statements to prove complex ones, not vice versa. We should use theories as fewer as possible and choose theories as simple as possible. Only through removing unnecessary theories from our argument may we make our solving process leaner and simpler.

(Discussion) When we solve a problem, we should avoid using unnecessary theories. After Guo–Wang [39, pp.62–63] discusses the Frobenius method in the general case, Guo–Wang [39, §4.4] repeats the same method many times. According to the general theory of regular singular points (Guo–Wang [39, §2.4 & §2.5]), shall we repeat the same discussion for the Legendre equation's three regular singular points given in Watson–Whittaker [88, p.304, l.18]? If we use a theory only because it is applicable, then our argument will become aimless. Fortunately, there are simpler methods available. Indeed, when we solve an ordinary differential equation, we should focus on directly finding a solution in closed form. We should not divert our attention to the solution's properties or generalization. See Example 3.29.

(h) Advantages: Take full advantage of available resources, circumstances and opportunities.

(i) Flexibility: Use the ideas in a theory flexibly rather than the exact form of theorems by mastering the entire theory. That is, we should apply the essential idea rather than the exact form of the general theorem to a specific case.

(j) Avoid contradictions: A theory cannot allow contradiction or inconsistency. If a theory cannot explain certain phenomena, we should modify it so that the new theory can explain them and in special cases the results of the new theory should reduce to those of the old one.

(k) Expand the scope of application without loss of efficiency

(3) Structurization

Theorems would become fragmented without structure; without σ-algebra, the statement given in Borovkov [13, p.47, l.8–l.9] would become fragmented [see Chung [17, p.65, Exercise 3]]. The critical structure may fail to emerge more often because of lacking in skilful analysis. In one varianle, the derivative of a function at a point is a number; in several variables, the differential of a function at a point is a matrix [Rudin [70, p.189, l.-12]]. For the limit-point case at ∞, we use a real nondecreasing spectral function [Coddington–Levinson [18, p.232, l.5]]; for the limit-point case at both $-\infty$ and ∞, we use a Hermitian nondecreasing spectral matrix [Coddington–Levinson [18, p.247; l.-17; l.-10;l.-7]]. The basic idea of differentiation of one variable and that of several variables are essentially the same, so are one-end and two-end limit-point cases [Coddington–Levinson [18, chap. 9, §3–§5]]. In the complicated case, the only thing we should pay attention to is the formation of a new structure–matrix. The concept of self-adjointness and eigenvalues in matrix theory can be used to classify the systems of differential equations [Coddington–Levinson [18, p.189, l.4]] and find their eigenfunctions [Coddington–Levinson [18, p.196, l.15–p.197, l.8]]. Consequently, structurization may help us systematically operate and deeply understand the complicated cases.

(a) Insightfulness: This quality refers to origins, insights, motivations, perspectives, true nature, inner structures, and formal solutions.

(Discussion) A proof should be well-structured and insightful. If the conclusion of a theorem is valid in most cases, then we simply apply the conclusion to a problem without checking if the situation satisfies the hypothesis of the theorem. Thus, we use the theorem first and justify the application latter. This formal procedure allows us to quickly obtain a solution candidate and have a crude blue print for solving the problem. In order to master the basic part of a subject, one should move ahead to study its advanced part.

(b) Essence: This quality refers to modeling, the common pattern of a solutions, key points, main veins, and the core of a theory. We should seek the common pattern of solutions in order to grasp the main vein that runs through the entire theory. Keen observations carry forward the method's development. If we use this main vein as the guideline to develop the theory, it may help us organize our material and clarify the theory. See Example 3.47 and Example 3.48. In order to avoid confusion and complexity, we must reduce various solution strategies to the essence. We should structuring the problem and locate the first obstacle to the solution so that we may easily and quickly recognize the reason why the problem cannot be solved with the assigned tools. The frequently used statements in a theory should be considered valued basics. Studying a complicated theory without understanding its essence is like returning from a treasure mountain with empty hands. The essence often becomes clearer if we reduce a complicated case to a simple case. Only through reducing a method to its essence may we be able to easily deal with complicated problems.

(c) Flowcharts or networks: This quality refers to relationships, compatibility, unifications, interactions, interdisplines, integration studies, external links, links among milestones, the big picture, flow charts in design, proof strategies, and the evaluation or criticism of a theory. A theorem is *inseprable* from its *role* in the entire theory. Only through advanced researches may we correct our mistakes in basics. A math network strengthens effectiveness. See Example 3.66.
(Discussion) Gauss, Lagrange, Kummer, and Hilbert wrote important work after mastering mathematics and physics. Their deep understanding of mathematical network enabled them to write masterpieces. It would be difficult to do so for those who specialize only a narrow field. Galois was able to write great work because he had read Lagrange's opus. Einstein had also read many people's work before he wrote papers on relativity. Mastering mathematical networks may help us deduce simplicity from apparent complicity, recognize the essence, understand the situation, propose important questions, and write significant papers.

A modern approach to a topic often focuses its study on a local, isolated and self-contained system. This approach will make it difficult to see the topic's origin and its role in the entire theory. Consequently, we should keep the external links open to help preserve the origin and the big picture. See Example 3.57 and Example 3.58.

The following examples indicate in parentheses the qualities of process for the methods under discussion.

Example 3.3. (Accessibility)

There are ghosts or no ghosts. This proposition is mathematically meaningless even though there are no other possibilities in logic.

Example 3.4. (Accessibility)

Suppose we want to prove the existence of the splitting field of a polynomial (Edwards [26, §51]). We must provide factorization methods which enable us to determine if a polynomial is reducible or irreducible in a finite number of steps (Edwards [26, p.69, l.24–l.28; p.72, l.15–l.19]). In order to avoid using unnecessary tools and resources, we consider only the following relevant statements:

(1) A polynomial with integer coefficients is either reducible or irreducible.
This statement was proved by Kronecker (Edwards [26, p.72, l.21–p.73, l.19]).

(2) A polynomial with rational coefficients is either reducible or irreducible (Edwards [26, p.73, Corollary 1]).

(3) Given a factorization method for the coefficient field K, one can find a factorization method for the coefficient field $K(a)$ obtained by adjoining to K an indeterminate a (Edwards [26, §58–§59]).

The key to proving the above statement is to reduce $f(a,x)$ with two variables to $\tilde{f}(t) = f(t^N,t)$ with one variable (Edwards [26, p.76, l.11–l.13]). This idea came also from Kronecker.

(4) Given a factorization method for the coefficient field K, one can find a factorization method for the coefficient field $K(a)$ obtained by adjoining to K a root a of an irreducible polynomial with coefficients in K (Edwards [26, §60]).

The key to proving the above statement is to use the method of undetermined coefficients by considering the factorization of the norm $Nf(x+ua)$ (Edwards [26, p.78, l.5 & l.10]).

Example 3.5. (Accessibility)
(Baire's category theorem) (Royden [69, p.139, Corollary 16]; Dugundji [24, p.251, Ex. 6])

The existence given in Dugundji [24, p.300, Theorem 4.2] is derived from reduction to absurdity. This existence is not as effective as the constructive existence given in Gelbaum–Olmsted [32, p.38, l.-6–p.39, l.3]. The fact that modern mathematicians rashly adopt short proofs (Royden [69, p.141, Exercise 30.d]; Rudin [72, p.121, Exercise 14]) but neglect effectiveness will reduce the quality of theory. Munkres [57, §7–§8] tries to improve the effectiveness of the existence given in Dugundji [24, p.300, Theorem 4.2]. The attempt is futile because Munkres' use of Baire's category theorem has seriously ruined effectiveness in the first place.

Example 3.6. (Accessibility)
(Constructing continuous functions that are non-differentiable)

In each of the sections in Titchmarsh [80, §11.21, §11.22 and §11.23], Titchmarsh constructs a continuous function that is not differentiable. The first one is simplest. This shows that we should start a project with a small task. The first and the third example show that if the derivative were to exist, it would have two different values. The second example shows that if the derivative were to exist, its value would be $+\infty$. Thus, the three constructions and proofs are similar. The reduction to absurdity Titchmarsh uses can be considered trivial. Thus, Titchmarsh's proofs are effective. In contrast, modern mathematicians love to use a non-trivial (see Example 3.5) reduction to absurdity to construct continuous functions that are non-differentiable. Due to their negligence the method of construction in modern textbooks deteriorates.

Example 3.7. (Accessibility; insightfulness; avoiding unnecessary complications)
(Cauchy's theorem)

In order to prove a theorem effectively, the quoted theorems in its proof should be simple, practical, and indispensable. The proof given in Rudin [72, p.221, Theorem 10.13] is more effective than those of Rudin [72, p.235, Theorem 10.35] and Saks–Zygmund [73, p.177, Theorem 2.3]. Rudin [72, p.224, Theorem 10.17] quoted in Rudin [72, p.236, l.-10] is impractical because in practice it is impossible to manage all the closed triangles in a open set. The use of difficult Runge's theorem in the proof of Saks–Zygmund [73, p.177, Theorem 2.3] obscures the essence of Cauchy's theorem.

Example 3.8. (Accessibility; directness; simplicity)

Both van der Waerden [82, vol.1, p.124, l.8–l.9] and Edwards [26, p.99, l.7–l.8] define the concept of a perfect field. The later definition is more accessible than the former one.

Example 3.9. (Accessibility; avoiding unnecessary complications)

Briefly speaking, Galois theory contains only two theorems: Edwards [26, p.59, l.7–l.11; p.61, l.-12–l.-8]. The former theorem builds only the Galois subgroup corresponding to an extension of the base field K obtained by adjoining the pth root of an element of K. The latter theorem builds only the subfield corresponding to a Galois subgroup whose index is a prime number p; this subfield is obtained by adjoining a pth root to the base field K; the proof of this theorem provides details about how we choose this pth root (Edwards [26, p.62, l.1–l.5; l.11–l.26]). In contrast, the proof given in van der Waerden [82, vol.1, p.156, the fundamental theorem] looks empty and impractical. For example, if Σ contains an infinite number of elements, it would be impossible to find the corresponding Galois subgroup in a finite number of steps. If we desire to effectively operate a subfield, our focus should be placed on the primitive element rather than all the elements of the subfield.

Example 3.10. (Reduction of calculations; quality of the output)

The method given in van der Waerden [82, vol.1, p.174, Lemma] is less effective than that given in Edwards [26, p.25, l.−8]. This is because the latter method produces an exact formula. For radicals, all the latter method requires is to take the 10th root of t^{10}, while the former method requires many extractions of roots. That is, the latter method requires less extractions of roots. This is the advantage of using Lagrange's resolvents.

Example 3.11. (Reduction to a lower-level system: reducing a partial differential equation to a system of ordinary differential equations)

The canonical Euler equations represent the characteristic system associated with the Hamilton–Jacobi equation [Fomin–Gelfand [30, p.90, l.−12–l.−10]].

Proof. Read Bendersky [7, p.182, l.−3–p.183, l.−1].
Bendersky [7, p.182, (56)] should be corrected as "$G(x_1,\cdots,x_m,u,p_1,\cdots,p_m)=0$, where $p_i \underset{\text{def}}{=} \frac{\partial u}{\partial x_i}$,".
Bendersky [7, p.183, (57)] follows from Tkachev [81, p.2, l.1–l.9]. □

Example 3.12. (Reduction with separation of variables in mind; reductions to lower order systems)

Given two differential equations Marion–Thornton [55, p.253, (7.87) & (7.88)]. We want to express λ in terms of θ using a *single* integration even though Marion–Thornton [55, p.253, (7.88)] is a differential equation of the *second* order. See Marion–Thornton [55, p.254, (7.93)]. The key to reducing the second order differential equation to the first order one is using Marion–Thornton [55, p.253, (7.90)]. Then we can solve the resulting differential equation by the method of separation of variables [Marion–Thornton [55, p.253, (7.91)]].
Remark. Hartman [40, p.50, l.−3–p.51, l.2, Lemma 3.1] provides an example of reduction by some known solutions.

Example 3.13. (Systematic reduction of calculations by using the formula for the derivative of a determinant)

In order to prove Watson [89, p.76, (7)–(11)], Watson suggests that we express successive derivatives of $J_\nu(z)$ and $Y_\nu(z)$ in terms of $J_\nu(z)$, $J'_\nu(z)$, and $Y_\nu(z)$, $Y'_\nu(z)$ by repeated differentiations of Bessel's equation. The method he suggests is not efficient for calculation. In fact, only the proof of (9) requires the differentiation of Bessel's equation. In order to derive the rest of formulae effectively and systematically, we should use the formula for the derivative of a determinant. For example, (7) follows from

$$\begin{vmatrix} J_\nu(z) & Y_\nu(z) \\ J''_\nu(z) & Y''_\nu(z) \end{vmatrix}' = \begin{vmatrix} J'_\nu(z) & Y'_\nu(z) \\ J''_\nu(z) & Y''_\nu(z) \end{vmatrix} + \begin{vmatrix} J_\nu(z) & Y_\nu(z) \\ J'''_\nu(z) & Y'''_\nu(z) \end{vmatrix} \text{ [Coddington–Levinson [18, p.28, l.16–l.17]].}$$

Example 3.14. (Analogy; reduction of calculations)

Watson–Whittaker [88, §15.8] discusses three theorems in the following order:

a The function given in Watson–Whittaker [88, p.329, l.21] satisfies the differential equation given in Watson–Whittaker [88, p.329, l.18].

b Watson–Whittaker [88, p.329, l.-11–l.-9, Theorem (I)]

c Watson–Whittaker [88, p.329, l.-7–l.-5, Theorem (II)]

In my opinion, following the above order is a bad approach. It is better if we discuss c first. This approach will enable us to quickly establish a relationship between $C_{n-r}^{r+1/2}$ and P_n^r. We can prove the equality given on Watson–Whittaker [88, p.329, l.-5] using Hobson [42, p.189, (12)] and Guo–Wang [39, p.276, (10)]. Without using c, it is difficult to prove a and b if one tries to follow the proof patterns given in Watson–Whittaker [88, p.303, l.-15–l.-9; p.304, l.6–l.13]. This is because $C_{n-r}^{r+1/2}$ differs from P_n^r by a factor containing $(z^2-1)^{-r/2}$. The difference will cause the number of terms to explore when we use differential operators (Watson–Whittaker [88, p.304, l.9]). It may also cause other problems when we compare coefficients (Watson–Whittaker [88, p.303, l.-13]). However, once c is proved, the proof of a and that of b will become easy. This is because there are *corresponding* properties of P_n^r ready for use. a follows from (c, Guo–Wang [39, p.250, (16)] and Watson–Whittaker [88, p.326, l.11]). b follows from (c, Guo–Wang [39, p.256, (8)] and Watson–Whittaker [88, p.324, l.18]). The formula given in Watson–Whittaker [88, p.324, l.18] is based on Ferrers' definition (Watson–Whittaker [88, p.323, l.-5]). When we apply the equality to the proof of b, we must add a factor of $(-1)^{1/4-\nu}$ to the right-hand side of the equality. This is because the notation P_n^r given in Watson–Whittaker [88, p.329, l.-5] is based on Hobson's definition (Watson–Whittaker [88, p.325, l.-3]) instead of Ferrers' definition (Watson–Whittaker [88, p.323, l.-5]).

Example 3.15. (Effectiveness: quantitative vs. qualitative formulations)
(The inverse function theorem)

Hartman [40, p.11, Exercise 2.3] provides a quantitative formulation about the inverse function theorem because it assigns the size of the ball D_1 on which f is one-to-one. In contrast, Rudin [70, p.193, Theorem 9.17] provides only a qualitative formulation because it is stated in terms of open sets. The former is a more effective formulation, because its solution is more informative.

Remark 1. (Proof of Hartman [40, p.11, Exercise 2.3]). By Hartman [40, p.10, Theorem 2.1], there exists a function $g : \{x : |x| \le b/M\} \to D$ such that $g \circ f = Id$. Similarly, there exists a function $h : y : |y| \le b/(MM_1) \to \tilde{D}_0$ such that $h \circ g = Id$. $h = h \circ (g \circ f) = (h \circ g) \circ f = f$ on $\{y : |y| \le b/(MM_1)\}$. It is unnecessary to use Hartman [40, p.5, Theorem 2.5].

Remark 2. The inverse function theorem provides a vantage point for us to see the insight of reason why uniqueness implies continuity (Coddington–Levinson [18, p.23, l.10–l.12]).

Example 3.16. (Effectiveness; avoiding unnecessary complications)

The use of a complicated theorem can make a constructive existence less effective. Sometimes, it is impossible for a generalized theorem to preserve the effectiveness of a specific case.
(The Riemann–Lebesgue Lemmas)

Suppose we want to use a computer to verify Watson–Whittaker [88, p.172, Theorem 9.41 (I)] for a given function. It is easier to effectively convert the argument to a computer program using the proof of Watson–Whittaker [88, p.172, Theorem 9.41 (I)] than it is to do so using the proof given in Rudin [72, p.109, l.-16–l.-6]. This is because a complicated theorem (Rudin [72, p.96, Theorem 4.25]) is used in Rudin [72, p.109, l.-13]. In addition, the method given in Rudin [72, §5.14] cannot be used to prove the specific case given in Watson–Whittaker [88, p.172, Theorem 9.41 (II)].

Example 3.17. (Effectiveness: how generalization affects effectiveness)

The concept of Galois resolvent can be generalized to that of a primitive element (Edwards [26, p.46, Exercise 13 & 14]). For each statement about the Galois resolvent in Edwards [26, §38–§41], we can replace the field generated by the Galois resolvent with the field generated by a primitive element. If we use the definition of the Galois group given in Edwards [26, p.51, l.10–l.23], we must prove the following statements: Edwards [26, p.53, l.-11–l.-8; p.51, l.-7–l.-6]. If we use the definition of the Galois group given in van der Waerden [82, vol.1, p.154, l.10 & l.24], we need not prove the above statements.

Example 3.18. (Effectiveness; take full advantage of available resources)

Through application we may gain a vantage-point for effectiveness because there are more resources available. Effectiveness is surely an ongoing trend for the development of mathematics.

We use reduction to absurdity to prove"path-connectedness $\Rightarrow$ connectedness" (Dugundji [24, p.115, Theorem 5.3]), so it is more straightforward and effective to check connectedness by finding a path between two points. Even though the formal structure for connectedness is firmly established (Chevally [16, p.36, Proposition 2]), we prefer proving connectedness by operating explicit paths (Fomenko [29, p.14, l.6; p.15, l.-2]) rather than by deducting from ineffective structure theorems (Chevally [16, p.37, l.2–l.4]).

Example 3.19. (Effectiveness; insightfulness: perspectives)

When we talk about effectiveness, we must know what aspect we refer to. Suppose we want to express a function as an infinite product. Sometimes, we refer to an entire process of constructing the expansion. If we want to construct the expansion *from scratch*, then Guo–Wang [39, p.25, Theorem 1] is more effective than González [37, p.202, Theorem 3.16]. However, if the identity for the expansion is already given (González [37, p.206, (3.5-7)]), we just want to prove its validity, then the less effective and more general theorem (González [37, p.202, Theorem 3.16]) can be the best choice. Compare the proof of González [37, p.206, (3.5-7)] with that of Guo–Wang [39, p.26, (2)].

Example 3.20. (Hitting multiple targets with one shot)

Guo–Wang [39, p.324, l.-2–p.325, l.7] finds the integrals given in Guo–Wang [39, p.325, (17)] with one shot, while the proof given in Watson–Whittaker [88, p.350, l.-11–p.351, l.12] is divided into two cases.

Example 3.21. (Hitting multiple targets with one shot)

Birkhoff uses Birkhoff–Rota [10, p.25, (23)] to simultaneously prove both the uniqueness and the continuous dependence on initial values.

Example 3.22. (Hitting multiple targets with one shot)

Hartman uses Hartman [40, p.9, l.-1] to prove the uniqueness and estimate the error term at the same time.

Example 3.23. (Hitting multiple targets with one shot)

Rudin [72, p.130, l.18].

Example 3.24. (Reduction of calculations; directness; simplicity)

It is simpler to use Ince [43, p.161, l.21–l.23] rather than Watson–Whittaker [88, p.202, l.16] to find out whether $z = \infty$ is a regular singular point of the second-order ODE.

Example 3.25. (Simplicity)
(The root test vs. the ratio test)

For convergence tests, we choose the ratio test for $\sum_{n=0}^{\infty} \frac{z^n}{n!}$, and choose the root test for $\sum_{n=0}^{\infty} (\frac{2n^2+1}{n^2+1})^n$.

Example 3.26. (Simplicity; directness; avoiding unnecessary complications)
(The Bessel function of order n)

The introduction of the Bessel function of order n given in Ince [43, p.189, l.-2–p.190, l.8] is simple and direct, while the introdution of the same function given in Watson–Whittaker [88, p.355, l.12–p.356, l.5] is unnecessarily complicated.

Example 3.27. (Simplicity)
(Recurrence relations for Bessel functions)

In order to prove the recurrence relations for Bessel functions, it is simpler to use the generating function (Watson–Whittaker [88, p.358, Example 1]) than use the integral representation (Watson–Whittaker [88, §17.21]).

The proofs given in Watson–Whittaker [88, §17.21] use integral representations, while those given in Guo–Wang [39, p.349, l.8–p.350, l.9] use series. The latter method is more elementary, so it is better.

Example 3.28. (Simplicity; directness)

Edwards [26, p.86, l.-7–p.87, l.10] discusses how Gauss and Kronecker proved the following statement:
$g(a)g(a^2)g(a^{p-1}) \equiv g(1)^{p-1} \pmod{p}$, where $a \neq 1$ and $a^p = 1$. Their ideas might be new during their times. Now we may adopt a simpler, more direct and inspiring method to prove the above statement using the following theorem:
If an ordinary integer is divisible by $a-1$, then it is divisible by p (Edwards [25, p.93, l.8–l.10]).

Example 3.29. (Simplicity; avoiding unnecessary complications)

Given that $P_n(x)$ is a solution of Lengendre's equation, find the second solution.

Solution. We should use the method given in Ince [43, p.166, l.15–p.167, l.7] rather than the one given in Guo–Wang [39, p.226, l.3–p.230, l.6] because the later method uses complicated Cauchy's integral formula (Guo–Wang [39, p.227, (5)]; Watson–Whittaker [88, p.303, l.-6]). We should quote Birkhoff–Rota [10, p.37, l.1–l.4] rather than Guo–Wang [39, p.63, (13)] to obtain the solution form given in Ince [43, p.166, l.17] because Guo–Wang [39, p.63, (13)] uses the complicated Frobenius method. □

Example 3.30. (Simplicity: using simple statements to prove complicated ones; avoiding unnecessary complications)

Both van der Waerden [82, vol.1, p.112, l.3–l.19] and Edwards [26, p.142, l.2–l.6] prove the existence of a primitive $(q-1)$st root of unity. The former approach is more direct and constructive because it uses simple statements to prove more complicated ones.

Example 3.31. (Simplicity; avoiding unnecessary complications)

In order to include all the solutions (Coddington–Levinson [18, p.115, l.-11–l.-10]), Levinson considers formal Laurent series (Coddington–Levinson [18, p.116, l.12]). In fact, through the transformation given in Hartman [40, p.79, (11.20)], it requires only the consideration of the formal power series (Hartman [40, p.80, l.14; p.78, Theorem 11.3]) rather than that of formal Laurent series (Coddington–Levinson [18, p.117, Theorem 3.1]). The complications given in Coddington–Levinson [18, p.115, l.-11–p.118, l.24] are unnecessary.

Example 3.32. (Advantages: taking full advantage of circumstances; reduction of calculations)

We should take advantage of the situation whenever possible. Given a polynomial P of degree n. We would like to express a symmetric polynomial of roots of P in terms of the coefficients of P (Jacobson [45,

vol. 1, p.109, Theorem 9]). For example, we want to find uv in terms of σ_1, σ_2, σ_3 (Edwards [26, p.22, Exercise 2]). First, we need a reference table such as the one given in Edwards [26, pp.6–7, (1)–(19)]. The method given in Edwards [26, pp.14–15, Exercise 12] applies to the general case. However, it would be ineffective to use this term by term transformation to solve our specific case. We should try to group terms and express them in terms of σ_1, σ_2, σ_3 whenever possible before we carry out our calculations. For example, the following observation saves us tremendous calculations: $(x^2y+y^2z+z^2x)^2+(y^2x+z^2y+x^2z)^2$
$=(x^2y+y^2z+z^2x+y^2x+z^2y+x^2z)^2-2(x^3y^3+y^3z^3+x^3z^3)-6x^2y^2z^2-2(xyz)(x^3+y^3+z^3)$.

Example 3.33. (Advantages: shortcuts; effectiveness; reduction of calculations)
$[\frac{d^{2p}}{d\theta^{2p}}\sum_{m=0}^{\infty}\frac{(-1)^m z^{2m}\sin^{2m}\theta}{(2m)!}]_{\theta=0}=[\frac{d^{2p}}{d\theta^{2p}}\sum_{m=0}^{p}\frac{(-1)^m z^{2m}\sin^{2m}\theta}{(2m)!}]_{\theta=0}$ [Watson [89, p.36, l.7–l.10]].

Proof. Because we only need to consider the θ^{2p} term, we may reduce the index set of summation from the infinite set $\{0,1,2,\cdots\}$ to the finite set $\{0,1,2,\cdots,p\}$. $\square$

Example 3.34. (Flexibility)

Levinson proves Coddington–Levinson [18, p.20, Theorem 5.1] using the exact form of the general theorem (Coddington–Levinson [18, p.12, Theorem 3.1]). However, the domain of the solution in the general case is not large enough to meet the requirement. Consequently, he uses Coddington–Levinson [18, p.15, Theorem 4.1] to extend the domain. In contrast, the proof of Pontryagin [65, p.167, l.-10–p.169, l.-1] applies the same idea directly to the entire domain. Consequently, it is immune from the fuss of domain expansion. The techniques of domain expansion given in Coddington–Levinson [18, p.15, Theorem 4.1] and Coddington–Levinson [18, p.20, Theorem 5.2] are unnecessary and insignificant.

Example 3.35. (Avoiding contradictions: from the Galilean invariance to the Lorentz invariance)

The Michelson–Morley experiment [Fowler [31]] suggests that the velocity of light be constant, independent of any relative motion of the source and the observer. However, the Galilean transformation is inconsistent with this suggestion [Marion–Thornton [55, p.548, l.-14–l.-7]]. Maxwell's equations are invariant in form under the Lorentz transformation [Wangsness [86, §29-5]]. For the contradiction caused by the Galilean invariance given in Marion–Thornton [55, p.549, l.9–l.11], the solution is to use Lorentz transformations instead of Galilean transformations. Muon decay provides an experimental verification of special relativity [Marion–Thornton [55, p.555, l.-14–p.556, l.-16]]. The speed of the light in invariant under Lorentz transformations [Marion–Thornton [55, p.552, l.1–l.5]]. Linear momentum is not conserved according to special relativity if we use the conventions for momentum of classical physics [Marion–Thornton [55, p.564, l.5–l.6]]. The key to solving this inconsistency is to use the *proper time* in the definition of linear momentum [Marion–Thornton [55, p.564, l.21–l.26; p.565, Example 14.6]]. Marion–Thornton [55, p.567, Example 14.7] shows that the relativistic kinetic energy reduces to the classical result for small speeds, $u \ll c$. If we use the position 4-vector $\mathbb{X}$ with $x_4 = ict$ to construct the Lorentz transformation matrix [Marion–Thornton [55, p.572, (14.77)]], then the momentum vector p given in Marion–Thornton [55, p.564, (14.45)] becomes the momentum–energy 4-vector $\mathbb{P}=(p, i\frac{E}{c})$ [Marion–Thornton [55, p.573, (14.91)]], where E is the total energy. The contradiction given in Marion–Thornton [55, p.574, l.-7–l.-1] forces us to modify the velocity addition rule: Marion–Thornton [55, p.576, (14.98)]. In order to make the Lagrange equations [Marion–Thornton [55, p.578, (14.107)]] accommodate the relativistic momentum vector [Marion–Thornton [55, p.578, (14.108)]], we must modify the definition of Lagrangian [Marion–Thornton [55, p.578, (14.113)]]. Because mass and energy are interrelated in relativity theory, it no longer is meaningful to speak of a "center-of-mass" system; in relativistic kinematics, one uses a "center-of-momentum" coordinate system instead. Such a system possesses the same essential property as the previously used center-of-mass system–the total linear momentum in the system is zero [Marion–Thornton [55, p.579, l.−16–l.−11]]. This modification of coordinate system leads to Marion–Thornton [55, p.582, (14.128); (14.129)] which are reduced to the classical results given in Marion–Thornton [55, p.350, (9.69); (9.73)] when $\gamma_1 \to 1$.

Example 3.36. (Expanding the scope of application without loss of efficiency: methods of determining the stability of circular orbits) [Marion–Thornton [55, §8.10]]

Marion–Thornton [55, p.317, l.5–l.7] gives a method of determining the stability of circular orbits. However, examples that may take advantage of this method are few. They are limited to simple functions such as that given in Marion–Thornton [55, p.317, (8.75)]. For the complicated function given in Marion–Thornton [55, p.319, (8.96)], the method would take a tremendous amount of calculations. Instead, we should use the method given in Marion–Thornton [55, p.317, l.19–p.319, l.−8].

Example 3.37. (Structurization: the critical structure may fail to emerge more often because of lacking in skilful analysis)

Coddington–Levinson [18, p.247, l.−9, (ii)].

Proof 1.

$$\begin{vmatrix} \sum_{m\le i\le n} r_{\delta i1}\bar{r}_{\delta i1} & \sum_{m\le i\le n} r_{\delta i1}\bar{r}_{\delta i2} \\ \sum_{m\le i\le n} r_{\delta i2}\bar{r}_{\delta i1} & \sum_{m\le i\le n} r_{\delta i2}\bar{r}_{\delta i2} \end{vmatrix}$$

$$\begin{aligned} &= \sum_{m\le i\le n}\sum_{m\le j\le n} r_{\delta i1}r_{\delta j2}\bar{r}_{\delta i1}\bar{r}_{\delta j2} - \sum_{m\le i\le n}\sum_{m\le j\le n} r_{\delta i1}r_{\delta j2}\bar{r}_{\delta i2}\bar{r}_{\delta j1} \\ &= \sum_{m\le i<j\le n} r_{\delta i1}r_{\delta j2}(\bar{r}_{\delta i1}\bar{r}_{\delta j2} - \bar{r}_{\delta i2}\bar{r}_{\delta j1}) + \sum_{m\le j<i\le n} r_{\delta i1}r_{\delta j2}(\bar{r}_{\delta i1}\bar{r}_{\delta j2} - \bar{r}_{\delta i2}\bar{r}_{\delta j1}) \\ &= \sum_{m\le i<j\le n} (r_{\delta i1}r_{\delta j2} - r_{\delta i2}r_{\delta j1})(\bar{r}_{\delta i1}\bar{r}_{\delta j2} - \bar{r}_{\delta i2}\bar{r}_{\delta j1}). \end{aligned}$$

□

Proof 2. Since the sum of positive semidefinite matrices is positive semidefinite, it suffices to observe

$$\begin{bmatrix} r_1\bar{r}_1 & r_1\bar{r}_2 \\ r_2\bar{r}_1 & r_1\bar{r}_2 \end{bmatrix} = \begin{bmatrix} r_1 & 0 \\ r_2 & 0 \end{bmatrix}\begin{bmatrix} \bar{r}_1 & \bar{r}_2 \\ 0 & 0 \end{bmatrix}.$$

□

Remark. The second proof can be generalized to prove Coddington–Levinson [18, p.263, l.4, (ii)].

Example 3.38. (Insightfulness)

Suppose we want to prove
$\int_0^\pi \frac{dx}{a+b\cos x} = \frac{\pi}{\sqrt{a^2-b^2}}(|b|<|a|)$.

Proof. Let $t = \tan\frac{x}{2}$.
$\frac{dt}{dx} = \frac{1+t^2}{2}$.
$\cos x = \frac{1-t^2}{1+t^2}$.
$\int \frac{dx}{a+b\cos x} = 2\int \frac{dt}{(a+b)+(a-b)t^2}$
$= \frac{2}{\sqrt{a^2-b^2}}\arctan(\sqrt{\frac{a-b}{a+b}}\tan\frac{x}{2}) + C$. □

The above proof by calculus is symbolic and incomplete because it is difficult to determine the value of the integral in some cases. Both Hobson [42, p.360, l.12–p.361, l.21] and Guo–Wang [39, p.268, l.2–l.16] prove the formula for $\int_0^{2\pi} \frac{dx}{A+B\cos x+C\sin x}$ in all cases. The former proof uses the integral of sin and that of cos, while the latter proof uses the residue theorem. In order to see why the statement given in Hobson [42, p.361, l.18–l.19] is true, we have to do some calculations from the viewpoint of the former proof. However, from the viewpoint of the latter proof, we can see the reason directly (Guo–Wang [39, p.268, l.12–l.16]). Consequently, the latter proof is well-structured and insightful.

Example 3.39. (Insightfulness; accessibility: narrowing the scope of mathematical induction; essence)

When we try to solve a problem, we should focus on the essence of the problem. Saks–Zygmund [73, p.108, Theorem 3.4] is a corollary of the general theorem given in Saks–Zygmund [73, p.107, Theorem 3.1]. The latter theorem specifies the conditions under which we may differentiate under the integral sign. Thus, the latter theorem highlights the essence of Saks–Zygmund [73, p.108, Theorem 3.4]. In contrast, Ahlfors [1, p.121, Lemma 3] provides only a trick to solve a particular problem. After studying Ahlfors' proof, we do not know the general method of solving this type of problem.

Another drawback of Ahlfors' proof is that the scope of mathematical induction used in his proof is too broad. We should limit the scope of mathematical induction used in a proof as narrowly as possible. In the proof of Saks–Zygmund [73, p.108, Theorem 3.4], the mathematical induction is used to derive $\frac{d[z^{(k+1)}]}{dz} = (k+1)z^{(k+2)}$. In contrast, Ahlfors tries to justify his differentiation in each induction step (Ahlfors [1, p.121, l.-8–p.122, l.8]). In other words, he differentiates under the integral sign countable times. On the one hand, it takes too much time figuring out trivial details; on the other hand, it consumes too much time and memory for computer if we allow too much work for each induction step.

Example 3.40. (Insightfulness: formal solutions)

In Guo–Wang [39, p.81, l.12], we assume that the interchange of the integral sign and L_z is valid. This procedure allows us to quickly obtain a formal solution $u(z)$ (Guo–Wang [39, p.81, (5) & (6); p.82, (7)]). The assumption will be justified later case by case. For example, in order to prove both that Guo–Wang [39, p.302, (2)] satisfies Guo–Wang [39, p.302, (1)] and that the integral given in Watson–Whittaker [88, p.339, l.-5] satisfies Watson–Whittaker [88, p.337, (B)], we use Rudin [72, p.27, Theorem 1.34] to justify the differentiation under the integral sign.

After we obtain a formal solution, it is easy to forget to prove it to be a true solution. In Guo–Wang [39, §6.4], Guo fails to rigorously prove that the formal solution given in Guo–Wang [39, p.302, (2)] is a solution of Guo–Wang [39, p.302, (1)]. Since the integral given in Guo–Wang [39, p.305, (1)] and the left-hand side of the equality given in Guo–Wang [39, p.305, (2)] are obtained by replacing zt in Guo–Wang [39, p.302, (2) and (3)] by $-t$, Guo–Wang [39, p.305, (1)] can only be considered a formal solution of the Whittaker equation (Guo–Wang [39, p.300, (1)]). Guo also fails to rigorously prove that this formal solution is indeed a solution. In contrast, Watson–Whittaker [88, p.339, l.-6–p.340, l.6] rigorously prove that the integral given in Watson–Whittaker [88, p.339, l.-5] is a solution of the Whittaker equation (Watson–Whittaker [88, p.337, (B)]). Note that Watson leaves out a factor, $(-1)^{-k-1/2+m}$, on the right-hand-side of the equality given in Watson–Whittaker [88, p.340, l.2–l.3].

Example 3.41. (Insightfulness: formal solutions)

We can quickly derive Guo–Wang [39, p.298, (6)] by replacing z in Guo–Wang [39, p.143, (10)] by z/β, and letting $\beta \to \infty$. This formal procedure is justified in Guo–Wang [39, p.302, l.2- p.303, l.5].

Example 3.42. (Insightfulness: formal solutions)

We can quickly derive Guo–Wang [39, p.303, (6)] from Guo–Wang [39, p.153, (7)] by interchanging α

and β, replacing z by z/β, and letting $\beta \to \infty$. This formal procedure is justified in Guo–Wang [39, p.303, l.10–l.17].

Example 3.43. (Insightfulness: perspectives)

Let $m \le [n/2]$. Then $\sum_{k=m}^{[n/2]} \binom{n}{2k}\binom{k}{m} = \frac{2^{n-2m-1} n[(n-m-1)(n-m-2)\cdots(n-2m+1)]}{m!}$.

Proof. Using the expansion of $(1+x)^n$ and then letting $x=1$ or -1, we can prove case $m=0$. Using the expansion of $\frac{d}{dx}(1+x)^n$ and then letting $x=1$ or -1, we can prove case $m=1$. However, it will become difficult to prove case $m \ge 2$ if we continue to use the above combinatory method. We shall resort to Bessel functions.

$(\sinh\theta + \cosh\theta)^n + (\sinh\theta - \cosh\theta)^n = \sum_{k=0}^{[n/2]} \sum_{r=0}^{k} \binom{n}{2k}\binom{k}{r} \sinh^{n-2k+2r}\theta.$

The result follows from Watson–Whittaker [88, p.375, l.18–l.20] and Watson [89, p.272, (4)]. The properties of Bessel functions match our needs naturally and perfectly just as the properties of the Riemann zeta function match the needs for proving the prime number theorem. □

Example 3.44. (Insightfulness)

Edwards [26, p.59, l.22–l.28] and van der Waerden [82, vol.1, p.158, l.1–l.8] both describe how the extension K' of the base field K affects the Galois group of a polynomial equation $f=0$. In contrast, the former approach is more insightful. If $f = g_1 \cdots g_k$ is a factorization of f into factors irreducible over K', the Galois group of $g_1 = 0$ given in van der Waerden [82, vol.1, p.158, l.1–l.8] considers only the roots of g_1 and loses connections with other roots of f (e.g., Edwards [26, p.60, l.−7–p.61, l.6; p.65, Exercise 4]).

Example 3.45. (Insightfulness: only through studying the advanced theory may we master the basic one)

One cannot master calculus unless one completes one's study in advanced calculus. One cannot master Coddington–Levinson [18, chap. 1] until he fully understands Coddington–Levinson [18, chap. 7, p.189, Theorem 2.1]. A topic represents merely a stage of a theory's development. If one understands every word in a textbook about a topic, it does not mean one masters the topic. This is because one understands the theory up to that topic, but has not applied the topic to the later part of the theory.

Example 3.46. (Essence; directness; simplicity)
(How Lagrange solved $Pp+Qq=R$ [Boole [11, p.318, l.2–p.319, l.11]])

Lagrange's original solution. $dz = pdx + qdy$.
$Pdz - Rdx = q(Pdy - Qdx)$.
Suppose $Pdz - Rdx = du$ and $Pdy - Qdx = dv$. Then $du = qdv$.
Since the left side of the equality is an exact differential, $q = \phi'(v)$, where $\phi'(v)$ is an arbitrary function of v.
By separation of variables, we have
$u = constant = \phi(v)$.
$du = 0 = \phi'(v)dv = 0$.
$(du = 0 = dv) \Leftrightarrow (\frac{dx}{P} = \frac{dy}{Q} = \frac{dz}{R})$.
Let the solutons of $\frac{dx}{P} = \frac{dy}{Q} = \frac{dz}{R}$ be solved for the constants of integration, thus
$u_1(x,y,z) = a, v_1(x,y,z) = b$ [Ince [43, p.47, l.−14]]. Then we identify u, v with the constants of integration a, b. □

Remark. Shall we assign u_1 to u or v? According to the above considerations, u and v are not symmetric. However, according to the analytic proof given in Sneddon [75, p.52, l.9–l.-9], u and v are symmetric [Sneddon [75, p.52, l.-12]]. Consequently, it doesn't matter which one u_1 should be asigned to. Sneddon [75, p.50, l.-10–p.52, l.7] gives a geometric proof about the equivalence between $Pp+Qq=R$ and $\frac{dx}{P}=\frac{dy}{Q}=\frac{dz}{R}$. However, among the three (original, analytic, and geomeric) proofs, Lagrange's original proof shows most clearly the key to solving $Pp+Qq=R$. Lagrange goes straight to the heart of the matter. In contrast, Sneddon seems to focus on its side problems; his approach may easily make readers unable to see the wood for the trees.

Example 3.47. (Essence: seeking the common pattern of solutions; accessibility)
(The Lagrange resolvent)

If we can solve problems for some specific cases, we would like to find a *common* method for the general case (Edwards [26, p.2, l.23–l.28]). According to Edwards [26, p.17, (1); p.20, l.4–l.7; p.21, l.2–l.5], Lagrange observed that all the roots can be expressed in terms of a resolvent and its conjugates. After analyzing his observations, he proved the theorem given in Edwards [26, p.33, l.-11; p.34, l.10–l.17] using Lagrange resolvents to solve a polynomial equation with low degree and multiple roots. The following path shows the evolution of Lagrange's resolvents: Edwards [26, (p.22, l.-16 & l.-13) $\to$ (p.25, (1)) $\to$ (p.29, l.-19)]; the right-hand side of the arrow is more organized than the left-hand side. Galois went further to associate the roots of a polynomial equation with the group of transformations that permute the roots. By creating the Galois resolvent t (Edwards [26, p.114, l.-10; p.119, l.-10–l.-7]), he was able to able to use it to *generate* the splitting field $K(a,b,c,\cdots)=K(t)$ of a polynomial equation $f(x)=0$, where $a,b,c,\cdots$ are roots of $f(x)=0$ (Edwards [26, p.114, l.11–l.22]). That is, he expressed the splitting field as a *simple* algebraic extension $K(t)$ of K.

Example 3.48. (Essence; insightfulness; accessibility; effectiveness: avoiding trial and error; networks: interactions)
(The Galois resolvent)

For the concepts of a group (Edwards [26, p.48, l.-12–l.-8; p.49, l.17–l.19]), a subgroup (Edwards [26, p.50, l.1–l.2]), or a normal subgroup (Edwards [26, p.50, l.17–l.20]), Galois approached them from two viewpoints: the viewpoint of the roots of the polynomial and the viewpoint of the group itself. The former viewpoint is natural, concrete, as well as insightful and provides an easy way to find the elements of the group. It allows interface. In contrast, the latter viewpoint is abstract, but provides effective methods to check whether these elements satisfy the conditions in a definition or prove the properties of a definition in an organized manner (Edwards [26, p.49, l.-22–l.-15; p.56, Exercise 1, 2, and 3]). In other words, the former viewpoint involves concrete transformations that permute roots. Consequently, we have concrete resources to work with. Nowadays, the definition of a group and that of a normal subgroup in most textbooks lack origins and resources, especially significant examples for students to gain hands-on experience for these concepts. When we try to solve a polynomial equation, we should directly work with the permutation group of roots rather than the general concepts of group theory. For example, although both Edwards [26, p.51, l.-22–l.-13] and van der Waerden [82, vol.1, p.154, l.7–l.13 & l.24] discuss the Galois group, the former discussion is more intuitive and to the point. This is because in Edwards [26, p.51, l.-22–l.-13] *the Galois resolvent* is used to define the Galois group *directly* in terms of conjugates (Edwards [26, p.51, l.18]), while in Jacobson [45, vol. 3, p.27, l.-8–l.-7] or van der Waerden [82, vol.1, p.154, l.7–l.13 & l.24] the Galois group is defined in terms of invariants (Jacobson [45, vol.3, p.27, l.15 & l.28–l.29]; van der Waerden [82, vol.1, p.154, l.9–l.10]). If the subfield has an infinite number of elements, then it is difficult to check the latter definition in a finite number of steps. Thus, the former definition is more accessible than the latter definition. Galois proved the latter definition as a theorem (Edwards [26, p.52, l.-10–l.-9]). The latter definition is useful in the generalization from the Galois group of a polynomial equation to the Galois group of a normal field with respect to the base field (van der Waerden [82, vol.1, p.154, l.10–l.11]).

Similarly, it is more natural, consistent (Edwards [26, p.84, l.−8–l.−4]), advantageous (Edwards [26, p.120, l.−23–l.−14]) to define a subgroup using the Galois resolvent than using group theory. It would make its meaning rich and its concept clear if we define a normal subgroup (Edwards [26, p.51, l.3–l.6]) using the Galois resolvent rather than group theory.

Remark 1. Although Lagrange's methods of selecting the roots of resolvent equation (Edwards [26, p.18, l.2–l.10; p.19, l.−2–p.20, l.7; p.21, l.6–l.19; p.126, l.−7–l.−4]) are marvelous, they are not always reliable. This is because they use the method of trial and error (Edwards [26, p.126, l.−6–l.−4]) and because one may easily get lost (Edwards [26, p.126, l.−4]) if there is no general guideline to follow. In order to fix this problem, one must read van der Waerden [82, vol.1, 169, l.16–l.−2; p.179, (2), (3), & (4)]. Especially, note the delicate design of the Lagrange resolvent given in van der Waerden [82, vol.1, p.169, (1)] and its useful formula (van der Waerden [82, vol.1, p.169, (3)]).

Remark 2. van der Waerden [82, vol.1, p.181, l.−4–l.−2] claims, "Each single Θ is fixed by 8 permutations; the three of them together remain fixed only under $\mathfrak{B}_4$." One may argue, "An automorphism σ fixing Θ_1 will fix Θ_2 because $\Theta_2 \in \Delta(\sqrt{D})[\Theta_1] = \Delta(\sqrt{D})(\Theta_1, \Theta_2, \Theta_3)$. This would contradict van der Waerden's claim."

Clarification to the above confusion. σ has the above property if σ belongs to the Galois group of $\Delta(\sqrt{D})[\Theta_1]/\Delta(\sqrt{D})$. However, the 8 permutations belong to the Galois group of Σ/Δ. If τ belongs to the Galois group of Σ/Δ, τ is determined only if its value at the the generator of Σ/Δ is determined. Θ_1 is the generator of $\Delta(\sqrt{D})[\Theta_1]/\Delta(\sqrt{D})$, but is not the generator of Σ/Δ.

Remark 3. To recognize the main vein of a theory requires talent, acumen, and mastery of related research (Edwards [26, p.2, l.11–l.17]).

Example 3.49. (Essence: proving insolvability by locating the first obstacle to solution; insightfulness)
(The general equation of degree $n > 4$ is not solvable by radicals)

In order to prove a problem is insolvable with assigned tools, we first proceed to find a solution until we meet obstacles. The lessons of failure may inspire us to approach the problem differently. In order to prove Edwards [26, p.91, l.7–l.8, Corollary], we must investigate the solutions of a cubic equation. The feature of the method given in Edwards [26, §14] is the use of Lagrange's resolvents. All the roots can be expressed in terms of Lagrange's resolvents. The organized solution given in Edwards [26, p.133, l.−9–p.134, l.−1] shows that if the Galois group G is solvable (Edwards [26, p.61, l.16]), we may divide the process of finding solutions into ν steps. Each step corresponds to a specific subgroup and thereby a specific subfield. In each step, Edwards [26, p.61, Proposition] provides a tool that we may use to proceed toward our goal. This structured solution indicates that if the Galois group is not solvable, we can locate the step in which we will encounter the first obstacle to the solution. Since the circumstances in the beginning of meeting obstacles is less complicated than those in later stages, it is easier for us to analyze and figure out the reason why the equation is not solvable. Intermediate field extensions were used by Gauss to simplify the construction of a pth root of unity (Edwards [26, p.29, l.1–l.15; p.30, l.12]). Galois' contribution was to find a subgroup corresponding to each subfield (Edwards [26, p.57, l.18–l.20]).
Remark. The same idea can be used to prove that it is impossible to solve the equation $x^3 + px + q = 0$ by *real* radicals for the case $D > 0$. See van der Waerden [82, vol.1, p.180, l.11–l.14].

Example 3.50. (Essence: finding the cases that Riccati's equation is integrable in finite terms; insightfulness)

If Riccati's equation $\frac{dy}{dx} = az^n + by^2$ is integrable in finite terms, then $n = -2$ or $-\frac{4m}{2m\pm1}$ $(m = 0, 1, 2, \cdots)$ [Watson [89, p.123, l.23–l.26]]. Since Riccati's equation is a variant [Ince [43, p.24, l.−11–p.25, l.4]; Watson [89, p.96, (6)]] of Bessel's equation, we may use the language of Bessel's functions to translate the above theorem as follows:

If Bessel's equation for functions of order ν [Watson [89, p.117, l.−10]] is soluble in finite terms, then 2ν is an odd integer.

In the proof [Watson [89, §4·7–§4·74]] of the latter version, despite possible difficulties, all the problems can be solved except one. That is, an infinite power series cannot be expressed as a polynomial [Watson [89, p.123, l.4]]. From this, we may determine the cases that Bessel's equation is soluble in finite terms [Watson [89, p.123, l.6–l.8]].

Example 3.51. (Essence: existence and uniqueness of solutions of ODEs)

There are 30 theorems and one corollary in Coddington–Levinson [18, chap.1 & chap.2]. They all discuss the existence and uniqueness of solutions of differential equations; many of them use the same method. Only through organization may we see essential solution strategies clearly.

(1) By treating parameters as initial values or treating initial values as parameters, we may reduce the number of problems from two to one.

By treating parameters as initial values (Coddington–Levinson [18, p.31, l.9–l.10]), Coddington–Levinson [18, p.30, Theorem 7.5] can be considered a simple application of Coddington–Levinson [18, p.25, Theorem 7.2]. Similarly, Coddington–Levinson [18, p.29, Theorem 7.4] can be considered an application of Coddington–Levinson [18, p.22, Theorem 7.1]. By treating initial values as parameters (Pontryagin [65, p.178, l.10–l.−3]), Pontryagin [65, p.179, Theorem 15] can be considered a simple application of Pontryagin [65, p.170, Theorem 13; p.173, Theorem 14; p.177, (B)].

(2) Generalization for solutions' continuity with respect to parameters:

Coddington–Levinson [18, p.29, Theorem 7.4]

→ **Coddington–Levinson [18, p.58, Theorem 4.1]**
The hypothesis is reduced from the Lipchitz condition to the uniqueness of solutions

→ **Coddington–Levinson [18, p.59, Theorem 4.2]**
In Coddington–Levinson [18, p.58, l.−5], we quote Rudin [72, p.27, Theorem 1.34] instead of Rudin [70, p.137, Theorem 7.14]

→ **Coddington–Levinson [18, p.59, Theorem 4.3]**
Replace φ_μ in Coddington–Levinson [18, p.58, l.15] with $\varphi_{\sigma,\eta,\mu}$

(3) Generalization for successive approximations:

The Lipschitz condition

Coddington–Levinson [18, chap.1, sec.3]

Continuity

Coddington–Levinson [18, p.4, Fig.1] : Using a polygonal line to approximate the flow of the differential equation

The generalized Lipschitz condition for Lebesgue integrals

Functional values in one section is determined by functional values of the previous section (Coddington–Levinson [18, p.43, (1.5)]); ODE is valid almost everywhere; solution is absolutely continuous; Coddington–Levinson [18, p.54, Theorem 3.1] asserts the convergence of successive approximations

(4) The proof of Coddington–Levinson [18, p.34, theorem 8.1] is the same as that of Coddington–Levinson [18, p.12, Theorem 3.1] except replacing the real case by the complex case and continuity by analyticity. The proof of Coddington–Levinson [18, p.22, Theorem 7.1] (Continuity of solutions with respect to initial values) is essentially the same as that of Coddington–Levinson [18, p.12, Theorem 3.1] except replacing Coddington–Levinson [18, p.12, (3.2)(i)] by Coddington–Levinson [18, p.24, (7.3)(i)] and $|t-\tau| \leq \alpha$ by V given on Coddington–Levinson [18, p.22, l.-3]. Coddington–Levinson [18, p.22, Theorem 7.1] proves $\varphi \in C$, while Coddington–Levinson [18, p.25, Theorem 7.2] proves $\varphi \in C^1$ by using Coddington–Levinson [18, p.26, (7.11)] and Coddington–Levinson [18, p.20, Theorem 5.1].

In brief, the successive approximations converges uniformly to the solution. The continuity and analyticity follow from the uniform convergence.

Example 3.52. (Essence: the frequently used statements in a theory should be considered valued basics)

Coddington–Levinson [18, p.244, (4.10); p.248, l.-6] use the construction given in Birkhoff–Rota [10, p.286, (67)], but the former book never expresses it as a theoremexplicitly.

Example 3.53. (Essence: the essence often becomes clearer if we reduce a complicated case to a simple case)

The proof of Sturm's oscillation theorem given in Coddington–Levinson [18, p.212, Theorem 2.1] is too complicated to read or understand its essence. The following steps may help recognize its essence:

(1) If we consider the solution of an equation of motion as the path of a particle, then the phase plane may offer the detailed behavior of the motion [Pontryagin [65, p.125, Figure 39–Figure 41] and Arnold [4, p.148, Fig. 106]]. If we use the Prüfer substitution, a concept similar to that of the Poincarè phase plane, the differential equation of the second can be reduced to that of the first order [Birkhoff–Rota [10, p.267, (22)]] and Birkhoff–Rota [10, p.27, Corollary 1] can be interpreted as Birkhoff–Rota [10, p.269, Lemma 1].

(2) Birkhoff formulates tree lemmas given in Birkhoff–Rota [10, pp.269–270] before he actually proves the Sturm oscillation theorem [Birkhoff–Rota [10, p.270, Theorem 4]].

If we are versed in the above basics, we will have no problem studying the proof of Coddington–Levinson [18, p.212, Theorem 2.1]. Now let us prove a few statements in this proof:

Statement 1. $\omega(t,\lambda) > k\pi$ for $t > t_k$ [Coddington–Levinson [18, p.212, l.17]].

Proof. Since $\omega'(t_k) > 0$, there exists an $\varepsilon > 0$ such that $\omega(t,\lambda) > k\pi$ on $(t_k+\varepsilon, t_{k+1})$. If $\omega(t,\lambda)(t \in (t_k, t_{k+1}))$ intersects with $\theta = k\pi$, this will contradict the definition of t_{k+1}. □

Statement 2. $\omega(c,-\lambda) \leq \delta$ for $-\lambda$ large enough [Coddington–Levinson [18, p.213, l.15]].

Proof. (a) If $\omega(c,-\lambda) > \pi - \delta$, there exists a t_1 such that $\omega(t_1,-\lambda) = \pi - \delta$. By Coddington–Levinson [18, p.213, l.12], $\omega(t)$ is decreasing at $t = t_1$. Once $\omega(t,-\lambda)(t_1 < t \leq c)$ intersects with $\theta = \pi - \delta$, it will starts to decrease. Hence, $\omega(t,-\lambda) \leq \pi - \delta$ on $(t_1, c]$.

(b) $(\omega(c,-\lambda) \geq 0$ [Coddington–Levinson [18, p.212, l.-6]] and $\omega(a,-\lambda) = \alpha < \pi - \delta) \Rightarrow \omega(c,-\lambda) - \omega(a,-\lambda) > -\pi$.

(c) Assume $\omega(c,-\lambda) > \delta$. Then $\omega(c,-\lambda) - \omega(a,-\lambda) = \omega(t_2,-\lambda)(c-a)$, where $t_2 \in (a,c)$. By 2a, $\omega(t_2,-\lambda) \leq \pi - \delta$. $\omega(t_2,-\lambda)$ cannot be less than δ. Otherwise, $\omega(t,-\lambda)(t \in (t_2,c])$ will intersect with $\theta = \delta$ at $t = t_3$. By Coddington–Levinson [18, p.213, l.13], $\omega(t)$ will start to decrease at t_3 and as t increases whenever $\omega(t)$ raises to δ, it will start to decrease. Hence, $\omega(t,-\lambda) \leq \delta$ on $(t_3, c]$. But this contradicts our asumption $\omega(c,-\lambda) > \delta$.

(d) By Coddington–Levinson [18, p.213, l.14], $\omega(c,-\lambda) - \omega(a,-\lambda) = \omega(t_2,-\lambda)(c-a) < -10$. This contradicts 2b.

□

Statement 3. $0 < \omega(t,\lambda_0) < \pi$ in (a,b) [Coddington–Levinson [18, p.213, l.-15]].

Proof. If $\beta = \pi$, then b will be the *first* zero of φ. In any case, $\varphi(t,\lambda_0)$ cannot have zeros in (a,b). Hence, $\omega(t,\lambda_0)$ can neither increase to π nor decrease to 0. □

Example 3.54. (Essence: only through reducing a method to its essence may we be able to easily deal with complicated problems)

Isoperimetric problems: The proof of Bendersky [7, p.143, Theorem 27.6] is simpler than that of Fomin–Gelfand [30, p.43, Theorem 1]. This is because the latter proof uses the complicated concept of variational derivative. See Fomin–Gelfand [30, p.28, l.-10; Figure 3]. Although the concept of variational derivative may help understand the circumstance, it does not have much to do with the purpose of calculus of variations. The calculus of variations uses y as a variable which is the basis of the theory and should not be related further to x. At most, the latter proof is only a special case of the former proof. Holonomic problems: Similarly, the proof Bendersky [7, pp.146–147, Theorem 27.8(i)] is simpler than that of Fomin–Gelfand [30, p.46, Theorem 2]. Non-holonomic problems: Bendersky [7, p.148, l.10–p.150, l.-5] gives a detailed proof of Fomin–Gelfand [30, p.48, Remark 1]; only through reducing a method to its essence may we be able to easily deal with complicated problems.

Example 3.55. (Networks: why we should emphasize integration studies)

It can be said that physics integrates various topics in differential equations. For example, Jackson [44, §3.11 Expansion of Green Functions in Cylinderical Coordinates] integrates the general form of 3-dim Green's function [Jackson [44, p.38, (1.31); p.125, (3.142)]], the 1-dim Green's function [Jackson [44, p.125, (3.143)]], the Sturm–Liouville system [Jackson [44, p.126, (3.145)]] and the Wronkian normalization [Jackson [44, p.126, (3.146)]]. In differential equations, we usually treat them as independent and disconnected topics. However, when we put them into one physical system simultaneously to serve a special purpose (for the present case, the computation of the potential of a unit point charge), we should consider and ensure the compatibility among these topics. They are interrelated. The assignment of the value of a parameter of one subsystem may affect another subsystem. By $(pW)' = \psi_1(p\psi_2')' - \psi_2(p\psi_1')'$, the Wronskian of two independent solutions of Jackson [44, p.126, (3.145)] is $c/p(x)$, where c is a constant. If we assign c to be -4π as in Jackson [44, p.126, (3.146)], we will obtain Jackson [44, p.125, (3.143)], which can be used to compute the potential of a unit point charge. If we use Birkhoff–Rota [10, p.286, l.20] instead, our calculation will not obtain the correct electric potential. Except for leading us to the consideration of compatibility, integration studies may also help us

(1). Trace back to natural origins. From the viewpoint of one dimension alone, the formula given in Birkhoff–Rota [10, p.286, l.20] looks artificial. However, once it combines with integration studies, it will become natural: the integration through the normalization of Jackson [44, p.126, (3.146)] reveals that the radial consequence Jackson [44, p.125, (3.143)] originates from the natural symmetric Green's function in three dimensions [Jackson [44, p.125, (3.142)]].

(2). Observe that a side problem for one subject may be the main problem of another subject. Coddington–Levinson [18, p.192, Theorem 2.2(iii)] says that as a function of t, G satisfies $Lx = lx$ for $t \neq \tau$. How about if $t = \tau$? Even though the answer may help better understand Coddington–Levinson [18, p.192, Theorem 2.2], we often ignore this side problem. This is because the δ-function is an indefinable object in the classical theory of ordinary differential equations. At best, we can only say that $G^{(n-1)}(\tau,\tau,l)$ does not exist [Coddington–Levinson [18, p.192, Theorem 2.2(ii)]]. However, in functional analysis, the δ-function can be rigorously defined [Rudin [71, p.141, l.−7–l.−3]]. Then the above side problem becomes interesting and can be completely solved [Rudin [71, p.206, Exercise 10; p.378, l.−6]].

Example 3.56. (Networks: flow charts in design, proof strategies; effectiveness: directness)

Given a task and available resources. By aiming at the goal, we may design a flow chart to complete our mission. Any digressive topic is meaningless for this project.

The general solution of

$\frac{d^2y}{dz^2} - \frac{\phi'(z)}{\phi(z)}\frac{dy}{dz} + [\frac{3}{4}\{\frac{\phi'(z)}{\phi(z)}\}^2 - \frac{1}{2}\frac{\phi''(z)}{\phi(z)} - \frac{3}{4}\{\frac{\psi''(z)}{\psi'(z)}\}^2 + \frac{1}{2}\frac{\psi'''(z)}{\psi'(z)} + \{\psi^2(z) - \nu^2 + \frac{1}{4}\}\{\frac{\psi'(z)}{\psi(z)}\}^2]y = 0$ is

$y = \sqrt{\frac{\phi(z)\psi(z)}{\psi'(z)}}\mathscr{C}_\nu\{\psi(z)\}$ [Watson [89, p.98, l.−15–l.−11]].

Proof. The proof strategy is to eliminate $\chi(z)$ in Watson [89, p.98, (12) & (13)].

$\chi(z) = \frac{\phi^{1/2}}{[\psi'(z)]^{1/2}[\psi(z)]^{\nu-1/2}}$ [Watson [89, p.98, (14)]].

Watson [89, p.98, (16)] follows by substituting $\chi(z)\{\psi(z)\}^\nu = \frac{\phi^{1/2}[\psi(z)]^{1/2}}{[\psi'(z)]^{1/2}}$ into Watson [89, p.98, (12)].

The strategy for expressing $\frac{\chi'(z)}{\chi(z)}$ in terms of $\phi(z)$ and $\psi(z)$:

$2\frac{\chi'(z)}{\chi(z)} = \frac{\phi'(z)}{\phi(z)} - \frac{\psi''(z)}{\psi'(z)} - (2\nu - 1)\frac{\psi'(z)}{\psi(z)}$ [Watson [89, p.98, l.11]].

The strategy for expressing $\frac{\chi''(z)}{\chi(z)}$ in terms of $\phi(z)$ and $\psi(z)$:

By differentiating the above equality with respect to z, we have

$\frac{\chi''(z)}{\chi(z)} = (\frac{\chi'(z)}{\chi(z)})^2 + \frac{1}{2}\frac{\phi''(z)}{\phi(z)} - \frac{1}{2}(\frac{\phi'(z)}{\phi(z)})^2 - \frac{1}{2}\frac{\psi'''(z)}{\psi'(z)} + \frac{1}{2}(\frac{\psi''(z)}{\psi'(z)})^2 - (\nu - \frac{1}{2})\frac{\psi''(z)}{\psi(z)} + (\nu - \frac{1}{2})(\frac{\psi'(z)}{\psi(z)})^2.$ □

Remark. We may, but need not, derive formulas for $\chi'(z)$ or $\frac{\chi'''(z)}{\chi(z)}$ from the expression for $\chi(z)$.

Example 3.57. (Networks: subgroups vs. subfields, theory vs. reality; insightfulness: perspectives [axiomatic approaches vs. heuristic approaches]; avoiding unnecessary complications; essence; accessibility) (Galois theory)

Both Edwards [26] and van der Waerden [82, vol.1, chap.5–chap.7] discuss Galois theory. The latter book uses axiomatization as its guideline to derive its important results. It asks what background is required in order to understand the essence of Galois theory. Then it designs a flow chart in logic that leads to various theorems of the theory. In contrast, Edwards [26] adopts a different approach. First, it explains clearly about the theory's origin. Second, it asks what tools are needed in order to solve the problem. Then it introduces the concepts of group, subgroup, and normal subgroup, and Galois group. Third, it asks what strategy should be used to solve the problem. The strategy is to use the theorem that a polynomial equation $f(x) = 0$ is solvable by radicals ⇔ Galois group is solvable (Edwards [26, p.61, l.17–l.19]). The ⇒ part provides an effective test (van der Waerden [82, vol.1, p.173, l.14]) for solvability because a finite group is easier to work with than a field. The ⇐ part sketches some guidelines about how to find the solutions (van der Waerden [82, vol.1, p.173, l.15]) if the Galois group of the equation is solvable. The key to proving this theorem is contained in Edwards [26, p.58, l.-3–l.-1]. Fourth, it shows how we use the theory to solve practical problems (Edwards [26, p.133, l.-9–p.135, l.-4; p.91, l.17–l.19]). The approach given in Edwards [26] enables us to distinguish important theorems (Edwards [26, p.34, l.10–l.11; p.64, Theorem]) from side theorems (Edwards [26, p.72, l.15–l.19; p.82, Exercises 10 & 11]).

Example 3.58. (Networks; simplicity; avoiding unnecessary complications; advantages: using qroup theory to discuss the splitting field of a polynomial)
(Definition of Galois groups)

Let ζ be a primitive 12^{th} root of unity. The Galois group of $x^{12} - 1 = 0$ over $\mathbb{Q}$ given in Edwards [26, p.51, (1)] considers all the roots of $x^{12} - 1$, i.e., the *factorization* of $x^{12} - 1$ into *linear factors*, while the Galois group of $x^{12} - 1 = 0$ over $\mathbb{Q}$ given in van der Waerden [82, vol.1, p.154, l.24] only focuses on the roots of $x^4 - x^2 + 1 = 0$. According to Edwards [26, p.94, Theorem] or van der Waerden [82, vol.1, p.154, l.8], the elements of the Galois group are $\zeta \to \zeta$, $\zeta \to \zeta^5$, $\zeta \to \zeta^7$, $\zeta \to \zeta^{11}$.

If we adopt the terminology of group theory, we can formulate the following theorems more naturally and precisely. Edwards [26, p.56, Exercise 2] states that a subgroup divides a group into right cosets. Edwards [26, p.120, l.22–l.25] states that a Galois subgroup divides a Galois group into right cosets. Edwards [26, p.122, l.6–l.8] states a subgroup divides a group into left cosets. The proof of the first and the third statements essentially use group theory alone, so the proofs are abstract and their resources are limited. In contrast, the proof of the second statement uses a Galois subgroup which corresponds to a specific subfield. The extra resources enrich the meaning of the proof through the interaction between subgroups and subfields. Edwards [26, p.64, l.-6–p.65, l.6] uses very awkward language to illustrate the following proposition:
$\frac{G_i(G_{i-1}\cap\tilde{G})}{G_i} \simeq \frac{G_{i-1}\cap\tilde{G}}{G_i\cap\tilde{G}}$.
It would be much clearer and simpler if we use the theorem given in van der Waerden [82, vol.1, p.141, l.-4] instead to prove this proposition.

Example 3.59. (Networks: recognizing a theorem's attributes helps find its proof and determine the role that it plays in a theory)

From the viewpoint of Hartman, Hartman [40, p.14, Corollary 3.1] and Hartman [40, p.14, Corollary

3.2] are corollaries of Hartman [40, pp.12–13, Theorem 3.1] because it is necessary to prove the existence of a maximum interval before discussing them. In my opinion, the two corollaries and the theorem are corollaries of Hartman [40, p.11, Corollary 2.1] because the key idea of the proofs of the former three is Hartman [40, p.11, Corollary 2.1].

Example 3.60. (Networks: a theorem with added features; simplicity)

A pizza with sausage topping is still a pizza; it will not become a steak. Hartman [40, pp.14–15, Theorem 3.2] is a generalization of Hartman [40, pp.4–5, Theorem 2.4] with the feature of maximum interval. Hartman [40, pp.12–13, Theorem 3.1] shows the existence of a maximum interval. Consequently, the proof of Hartman [40, pp.14–15, Theorem 3.2] is the proof of Hartman [40, pp.4–5, Theorem 2.4] plus that of Hartman [40, pp.12–13, Theorem 3.1], and nothing else. If we fail to relate Hartman [40, pp.14–15, Theorem 3.2] to other theorems, its proof could look quite complicated.

Example 3.61. (Networks: the scope of a method's applicability)

When a method is related to a problem, we should apply the method to only where it may, and leave the rest to be dealt with in another way. In a finite-dimensional normed space, its various norms are equivalent (Rudin [71, pp.14–15, §1.19]). Consequently, all the properties of finite-dimensional normed spaces remain valid if we replace one norm with another. However, we cannot use this method to prove that Hartman [40, p.26, Lemma 3.2] implies Hartman [40, p.26, Exercise 3.1]. Instead, we should prove the latter statement as follows:

Proof. Let $h > 0$.
$\lim_{h\to 0} \frac{|y^j(t+h)|-|y^j(t)|}{h} = \frac{dy^j}{dt} \operatorname{sgn} y^j(t) (j = 1, \cdots, d)$ (Hartman [40, p.26, Lemma 3.1]).
$\lim_{h\to 0} \frac{(|y^1(t+h)|-|y^1(t)|, \cdots, |y^d(t+h)|-|y^d(t)|)}{h} = (\frac{dy^1}{dt} \operatorname{sgn} y^1(t), \cdots, \frac{dy^d}{dt} \operatorname{sgn} y^d(t))$.
Taking the Euclidean norm on both sides, we have $D_R|y(t)| = |y'(t)|$. □

Example 3.62. (Networks: inseparability of a theorem from its role in the entire theory)

We often isolate certain facts from a context and give them the status of a theorem. In doing so, we ignore the inseparability of a theorem from the role it plays in the entire theory.

For the application of a theorem, we would like to study how often it appears in the theory, in which areas it appears, and in what form it fits into its surroundings. For example, let us compare how John [46, p.153, l.8] and Rudin [71, p.180, l.10] introduce the Paley–Wiener theorem into the theory of PDE.

Example 3.63. (Networks: relationships)

We may use Green's functions, integral transforms, or separation of variables to solve PDE's (Sneddon [75, chap.3]). However, these methods are closely related. More precisely, each method *brings out* the next one.

We need *Green's function* to solve Sneddon [75, p.294, (1),(2),(3)].
→ We use *the Laplace transform* to determine the Green function (Sneddon [75, p.297, l.2]).
→ In view of the expansion given in Sneddon [75, p.298, (16)], we may also use *separation of variables* to determine the Green function (Sneddon [75, p.298, l.6]).

Example 3.64. (Networks: going back to the *basics* to establish the *main relationship*, unifications)

(1) The lack of development may make us mistake a partial aspect for the big picture.

The way Goldstein uses the inertia tensor may make his readers believe that tensors and linear transformations are the same (Goldstein [35, p.147, l.14–l.29]). In fact, this viewpoint can be justified

only under certain conditions (Warner [87, p.55, (d)]).
Note. Goldstein [35, p.146, (5-9)] is a corollary of Warner [87, p.55, (e)].

(2) The relationship between the basic concepts (the tensor algebra and the exterior algebra) can be fully established (Warner [87, p.56, Definition 2.4]).

The tensor product and wedge product in Warner [87, pp. 54–65] establish the relationships among three isolated major treatments: Lang's algebraic treatment (Lang [50, chap. XVI]), Goldstein's treatment for mechanics (Goldstein [35, p.146, l.15–p.147, l.−8]), and Spivak's superficial treatment for differential geometry (Spivak [77, vol.1, chap. 4 & chap. 7]).

   (a) The domains of the tensor product and the wedge product are fully developed.
   (b) The operations become simple and direct. Example (tensor product): Compare Warner [87, p.54, Definition 2.1] with Spivak [77, vol.1, p.159, l.7].
   (c) The artificial outlook of synthetic properties can be illustrated by inner basic operations.
   Compare Warner [87, p.56, Definition 2.4; p.57, 2.6(a)] with O'Neill [59, p.153, l.−1].

(3) (Inclusiveness and consistency)
The scheme in Warner [87, pp.54–62, §2.1–§2.13] is consistent with almost every existing concept of product.

   (a) Multiplication of real numbers (Warner [87, p.59, l.14]).
   (b) Scalar product (Arnold [3, p.173, Problem 7]).
   (c) Vector product (Arnold [3, p.173, Problem 6]).
   (d) (Lie group with its Lie algebra)
   Quaternion product = vector product − scalar product (Pontryagin [66, p.170, l.1]).
   Bracket product = vector product (Pontryagin [66, p.384, Example 93]).

(4) (Sorting)
We would like to distinguish between the identification by general properties (the 2nd isomorphism in Warner [87, p.60, l.2]) and the identification by a particular assignment (the 1st isomorphism in Warner [87, p.60, l.2]). Only for the latter may we have the *freedom* to make a choice for adjustment (Warner [87, p.60, l.5]).

(5) The basics are developed in order to study advanced topics that require clarification. Most common mistakes committed by mathematicians are basics. These basics may look confusing unless they are well-isolated from complicated situations.

(6) The basics are the foundation in building and expanding a theory. They are constantly modified by experimental results in order to advance further research.

Example 3.65. (Networks: quality checklist for a theory of tensors)

(1) Does the theory distinguish a *bound* vector from a free vector? Good: Kreyszig [49, p.103, l.12–l.16]. Poor: Peebles [63, §8].

(2) Does the theory mention that the allowable coordinate transformations form a *group*? Good: Kreyszig [49, p.101, l.20–l.21]. Poor: Peebles [63, §8].

(3) Does the theory have a *clear* definition of a tensor *field*? Good: Kreyszig [49, p.111, l.12]. Poor: Peebles [63, §8].

(4) Does the theory have a *consistent scheme* for development? Kreyszig [49, (31.1), (31.2), (31.3) & (32.1)] are proved according to the same scheme, while Peebles [63, p.230, (8.14)] is stipulated by hard and fast rules.

(5) Does the theory have a *geometric interpretation* for the contravariant or covariant components of a vector? Good: Kreyszig [49, p.116, Fig. 35.1 & Fig. 35.2]. Poor: Peebles [63, §8].

(6) When we use the elements of a vector space as contravariant vectors (Kreyszig [49, p.121, l.-11–l.-10]) and the elements of its dual as covariant vectors (Kreyszig [49, p.123, l.10]) to define tensors, do we relate it to the classical definition with a proper justification? Good: Kreyszig [49, p.122, l.9; p.123, l.-7–l.-6]. Poor: Spivak [77, vol. 1, chap. 4].

Example 3.66. (Networks: how a math network strengthens effectiveness)
(Identification of fundamental groups)

(1) Product: Pontryagin [66, p.350, F)].

(2) Homomorphic image: Pontryagin [66, p.370, l.-7]. A covering space is a generalization of a topological group homomorphism (Pontryagin [66, p.134, C)]). The use of universal coverings makes it easy to identify the fundamental group of a homomorphic image.

(3) Although we can define the fundamental group in an arcwise connected topological space R (Pontryagin [66, p.348, Definition 44]), for its identification we often wonder where to start if R does not have any group structure.

Remarks. Massey [56] completely ignores the role of topological group homomorphisms when discussing covering spaces. Pontryagin [66] shows the complete development process: Topological *group* homomorphism $\to$ covering space $\to$ covering *group*. Imposing a group structure on a covering space is like going back to the original stage. Warner [87] lacks the first part of the development process (Topological group homomorphisms $\to$ covering space).

Example 3.67. (Networks: links among milestones)

The correct approach toward developing a theory is to use the important facts as milestones and then link these facts with theorems. In contrast, the incorrect approach is to use the big theorems as the milestones and link them with examples.

In Courant–John [21, vol. 1, p.359, l.-7], the evolute E of a curve C is defined as the locus of the centers of curvature of C. In Courant–John [21, vol. 2, p.301, l.5], the evolute E of a curve C is defined as the envelope of the normals of C. In Courant–John [21, vol. 1, p.424], Courant proves that the first definition implies the second one. In Courant–John [21, vol. 2, p.301, Example 11], Courant proves that the second definition implies the first one. Weatherburn [90, vol. 1, §10 & §11] discuss involutes and evolutes, but fail to link evolutes with the concept of centers of curvature.

4 Methods of weakening a hypothesis

Suppose we use input-process-output as a model for a method. In some cases, we need to increase the input as shown in the definition of productiveness. In other cases, we need to reduce input. First, a theory is less likely to be contradictory and more likely to be consistent if it contains fewer assumptions.

Example 4.1. (Reducing the input for consistency; hitting multiple targets with one shot)

The four formulas given in Born–Wolf [12, p.41, (19)] are derived from a single figure: Born–Wolf [12, p.39, Fig. 1.10]. In contrast, the two formula given in Jackson [44, p.305, (7.38)] are derived from Jackson [44, p.305, Fig. 7.6(a)]. The two formulas given in Jackson [44, p.306, (7.40)] are derived from Jackson [44, p.305, Fig. 7.6(b)]. It can be said that Born's method hits two birds with one stone. Born–Wolf [12, p.39, Fig. 1.10] involves only one convention: Born–Wolf [12, p.40, l.−12–l.−11]. All the calculations in proving Born–Wolf [12, p.41, (19)] follow this convention. Jackson [44, p.305, Fig. 7.6(a)] and Jackson [44, p.305, Fig. 7.6(b)] involve two different conventions: Wangsness [86, p.411, l.−12–l.−11; p.415, l.4–l.5]. More conventions only increase the chance of leading to a contradiction. The explanations given in Wangsness [86, p.416, l.−12], Jackson [44, p.306, l.7] and Born–Wolf [12, p.42, l.−3–l.−1] are not satisfactory. In my opinion, for normal incidence, we should consider E parallel to the plane of incidence because this way fixes the value of E. If we were to consider E perpendicular to the plane of incidence, it would be difficult to determine whether E is positive or negative. The dichotomy given in Jackson [44, p.305, Fig. 7.6] unnecessarily uses the same set formulas (Jackson [44, p.304, (7.37)]) twice and fails to produce any extra benefit.

Second, the goal of axiomatization is to minimize a set of axioms that deduce the entire theory. If $A \Rightarrow B$, then we say that A is stronger than B or that B is weaker than A. Although axioms are stronest statements in a theory, we want to minimize the number of axioms. Third, in order to characterize a concept, we should seek a minimal set of its necessary conditions strong enough to become its sufficient conditions.

Example 4.2. (Reducing the input for characterizing a concept)

In the proof of Perron [64, p.276, Satz 38], Perron shows that the sufficient conditions for convergence originate from its necessary conditions (Perron [64, p.274, l.−6–l.−4]). This approach enables us to see how the theorem is produced and formulated. In contrast, the proof given in Wall [83, p.37, Theorem 8.1] fails to explain how the sufficient conditions are obtained because it fails to provide a reason for the artificial classification of cases. See Wall [83, p.38, l.4–l.5; l.17].

Fourth, in order to make a theorem stronger, we should weaken its hypothesis while keeping its conclusion the same. The goal of using the method of weakening a hypothesis is to find the weakest hypothesis for a given conclusion. Suppose A is the hypothesis and B is the conclusion of a theorem. If $A_1 \Rightarrow A_2 \Rightarrow A_3 \Rightarrow \cdots \Rightarrow A_n \Rightarrow B$, we want to find A_n, where the n is the largest. That is, we want to shorten the deduction chain. If Theorem A and Theorem B have the same conclusion and the hypothesis of Theorem A is weaker than that of Theorem B, then we may use a proof of Theorem A to prove Theorem B even though it may not be the most effective method to prove Theorem B. We say that the most effective proof of Theorem A is more refined than the most effective proof of Theorem B if Theorem A and Theorem B have the same conclusion and the hypothesis of Theorem A is weaker than that of Theorem B. If a hypothesis is modified so that it can be applied to a wider class, then the hypothesis is considered weakened. By weakening the hypothesis of a theorem, we may pinpoint the exact reason that leads to the conclusion. In the rest of this section, we will discuss methods of weakening the hypothesis alone.

4.1 Examples of weakening the hypothesis of a theorem while keeping its conclusion the same

In the following chains, the hypothesis of each theorem is weaker than that of the previous theorem:

(1) Zygmund [92, vol.1, p.78, l.-7–l.-6] $\to$ Zygmund [92, vol.1, p.78, Theorem 1.26]
Hypothesis: $(u_\nu = o(1/\nu)) \to (u_\nu = O(1/\nu))$

(2) Zygmund [92, vol.1, p.81, Theorem 1.36] $\to$ Zygmund [92, vol.1, p.81, Theorem 1.38]
Hypothesis: $(u_n = o(1/n)) \to (u_n = O(1/n))$

(3) Zygmund [92, vol.1, p.89, Theorem 3.4] $\to$ Zygmund [92, vol.1, p.90, Theorem 3.9]
Hypothesis: (x_0 is a point of continuity of f) $\to$ [$\Phi_{x_0}(h) = o(h)$ (Zygmund [92, vol.1, p.50, l.13; p.65, l.-12])]

(4) (Cahcuy's integral theorem: Conway [20, p.73, Proposition 2.15])(a consequence of Green's theorem) $\to$ Rudin [72, p.221, Theorem 10.13]
Hypothesis: analyticity (Conway [20, p.34, Definition 2.3]) $\to$ differentiability (Conway [20, p.96, Goursats Theorem])

(5) The Stone-Weierstrass theorem: Rudin [70, p.146, Theorem 7.24](the real case [resppectively, the complex case]) $\to$ Rudin [70, p.150, Theorem 7.30] [respectively, Rudin [70, p.152, Theorem 7.31]]
Hypothesis: ($\mathscr{A}$ is the algebra of real [respectively, complex] polynomials) $\to$ ($\mathscr{A}$ satisfies the hypothesis of Rudin [70, p.150, Theorem 7.30][respectively, Rudin [70, p.152, Theorem 7.31]])

(6) Uniqueness theorems about generalized Lipschitz conditions: Coddington–Levinson [18, p.10, Theorem 2.2] $\to$ Coddington–Levinson [18, pp.48–49, Theorem 2.1] $\to$ Coddington–Levinson [18, p.49, Theorem 2.2] (respectively, Coddington–Levinson [18, p.51, Theorem 2.3])
The hypothesis of Coddington–Levinson [18, pp.48–49, Theorem 2.1] is weaker than that of Coddington–Levinson [18, p.10, Theorem 2.2] (see Coddington–Levinson [18, p.49, l.12–l.19]). The hypothesis of Coddington–Levinson [18, p.49, Theorem 2.2] (respectively, Coddington–Levinson [18, p.51, Theorem 2.3]) is weaker than that of Coddington–Levinson [18, pp.48–49, Theorem 2.1] [see Coddington–Levinson [18, p.49, l.20] (respectively, Coddington–Levinson [18, p.51, l.-2])].

Remark 1. There can be following two versions of Zygmund [92, vol.1, p.90, Theorem 3.9]:

A $\sigma_n(x) \to f(x)$ for every x satisfying $\Phi_x(h) = o(h)$.

B $\sigma_n(x) \to f(x)$ almost everywhere.

If we adopt version 1A, we may use it to prove Zygmund [92, vol.1, p.89, Theorem 3.4]. In contrast, if we adopt version 1B, we will reach a point of no return. Namely, we can no longer use version 1B to prove Zygmund [92, vol.1, p.89, Theorem 3.4]. This is because the existence of x in version 1A is constructive (more specifically, x is fixed), while the existence of x in version 1B is less effective because it is derived from reduction to absurdity. Modern mathematicians love to use the term "almost everywhere" in real analysis simply because the meaning of this term is easier to remember than the meaning of $\Phi_x(h)$. This is the reason why delicate methods of weakening a hypothesis have almost become endangered species in real analysis.

Remark 2. Let $T = \{z \in C||z| = 1\}$, $e^{i\theta_0} \in T$ and $U \in L^1(T)$. (If $U \in C(T)$, then $P_U(z)$ [Ahlfors [1, p.167, l.$-$14]] is continuous on $\{z \in C||z| \leq 1\}$) $\rightarrow$ [If U is continuous at $z = e^{i\theta_0}$, then $\lim_{z\to e^{\theta_0}} P_U(z) = U(e^{i\theta_0})$ (Ahlfors [1, p.168, Theorem 25])]. See Ahlfors [1, p.167, l.9–l.12] for motivation. In contrast, the use of the phrase “almost everywhere” in Rudin [72, p.258, Corollary] prevents us from knowing the *exact* locations of $z = e^{i\theta}$ at which the formula given in Rudin [72, p.258, l.4] is valid.

4.2 How we recognize and appreciate the value of methods of weakening a hypothesis

In order to fully understand a method of weakening the hypothesis, we should not only know what it is, but also recognize its value and key points.

(1) We want to know from where the method comes. What problems motivate mathematicians to create such a device? What obstacle does this method of weakening a hypothesis can conquer, while other old methods cannot?

Suppose $z = \infty$ is a singularity of the second kind, we know the solutions of Coddington–Levinson [18, p.151, (4.1)] for the real case, and we want to find the solutions for the complex case (Coddington–Levinson [18, p.161, l.$-$8–l.$-$5]). Then it requires to replace the boundedness of f at $z = \infty$ in Conway [20, p.125, Theorem 1.4] with a growth condition, i.e., to prove Conway [20, p.135, Corollary 4.2]. See Conway [20, p.124, l.$-$2–p.125, l.1] and Coddington–Levinson [18, p.164, l.$-$10].

(2) We should not take a musket to kill a butterfly

We should highlight the amazing effects that a refined method of weakening the hypothesis produces. If an old, crude method can do, it is unnecessary to use a new, refined method of weakening the hypothesis. Using refined methods to do crude things is a unnecessary waste. For example, it is unnecessary to use the Phragmen–Lindelöf method to prove Rudin [72, p.274, Theorem 12.8]: we can prove the statement given in Rudin [72, p.275, l.11] using Conway [20, p.125, Theorem 1.4]. To specify a bound given the boundedness of f at $z = \infty$ is not as amazing as to specify a bound given the growth condition of f because the condition of the latter statement is weakened. Compare Rudin [72, p.274, Theorem 12.8] with Conway [20, p.135, Corollary 4.2].

(3) How to highlight the key idea of a method of weakening the hypothesis

(a) Use the method of standardization to eliminate unnecessary complications. For example, use a symmetric case (Conway [20, p.135, Corollary 4.2]) to represent the general case (Coddington–Levinson [18, p.162, Theorem A]) without loss of generality. See Conway [20, p.135, l.$-$3–l.$-$1].

(b) For the formulation of a method of weakening the hypothesis, we should trace the method’s origin and preserve its original setting. For example, Conway [20, p.135, Corollary 4.2] is a right version; see Conway [20, p.124, l.$-$1–p.125, l.1]. Adopting other versions such as Conway [20, pp.134–135, Theorem 4.1] or Rudin [72, p.276, Theorem 12.9] may distract us from the essence of the Pragmen–Lindelöf method.

5 Physical methods

5.1 Physical interpretations of a *problem*

Consider Laplace's equation (Watson–Whittaker [88, p.386, (I)] and Born–Wolf [12, p.11, (7)]). When we select a *coordinate system*, we should choose one suitable for the geometric *symmetry* of the shape of object (Jackson [44, p.104, l.6–l.12]).

5.2 Physical interpretations of a *solution*

Consider the solutions given by Born–Wolf [12, p.16, (8)] and Watson–Whittaker [88, p.397, l.8–l.19].

(1) Physical considerations help select meaningful solutions (Jackson [44, p.107, l.-2–l.-1] and Cohen-Tannoudji–Diu–Laloë [19, p.648, (C-9); p.652, l.17, p.664, l.2]).

(2) Solutions must be *well-defined*: In Jackson [44, p.104, l.15], we consider $x(= \cos\theta)$ instead of θ; in Jackson [44, p.105, l.-5], we restrict r to be great than 0.

(3) Physical considerations help select an appropriate solution *form* (Jackson [44, p.104, l.-6–l.-5]).

Example 5.1. (Solving the problem of coupled harmonic oscillators: Marion–Thornton [55, chap. 12]; Cohen-Tannoudji–Diu–Laloë [19, vol.1, pp.575–585, Complement H_V])

(1) From the viewpoint of differential equations: by changing variables [Marion–Thornton [55, p.471, (12.11)]], we may make the coupled differential equations given in Marion–Thornton [55, p.470, (12.1)] completely separable [Marion–Thornton [55, p.471, (12.14)]].

(2) From the viewpoint of individual particles using the Newtonian mechanics: The results are summarized in Marion–Thornton [55, p.487, Table 12-1]; the pictorial features are given in Marion–Thornton [55, p.472, Fig. 12-2].

(3) From the viewpoint of the entire system using the Lagrangian in the Lagrangian mechanics:

   (a) If the equations connecting the generalized coordinates and the rectangular coordinates do not explicitly contain the time, then the kinetic energy has the form given in Marion–Thornton [55, p.476, (12.18)].

   (b) The expansion of the potential energy in a Taylor series about the equilibrium configuration yields Marion–Thornton [55, p.476, (12.32)].

   (c) The Lagrangian equations yield Marion–Thornton [55, p.478, (12.38)]. By substituting Marion–Thornton [55, p.478, (12.39)] into Marion–Thornton [55, p.478, (12.38)], we have Marion–Thornton [55, p.479, (12.40)]. In order to find the solutions of Marion–Thornton [55, p.479, (12.40)], we solve ω for Marion–Thornton [55, p.479, (12.42)] first. Then for each ω_r, we solve Marion–Thornton [55, p.479, (12.40)] to obtain the corresponding eigenvector a_r.

   (d) Using Marion–Thornton [55, p.483, (12.63)], we simultaneously diagonalize T and U [Marion–Thornton [55, p.484, (12.65); (12.66)]]. Then the Langrangian equations in normal coordinates become completely separable [Marion–Thornton [55, p.485, l.4]].

(4) From the viewpoint of the entire system using the Hamiltonian operator in quantum mechanics: The first equality given in Marion–Thornton [55, p.480, (12.45)] is a special case of Cohen-Tannoudji–Diu–Laloë [19, vol.1, p.576, (4)]. By Cohen-Tannoudji–Diu–Laloë [19, vol.1, pp.584–585, Complement H_V, 2d]], we find that $< X_G > (t)$ and $< X_R > (t)$ oscillate at angular frequencies of ω_G and ω_R, which agrees with the classical result.

Remark. As we go to a more advanced level and widen our consideration, new physical meanings of mathematical equations continue to develop and meanings of equations become richer and more delicate. Nonetheless the meanings in older theories are still well-preserved in a newer theory.

5.3 How we understand the physical meaning of a mathematical theorem

In view of Jackson [44, p.36, l.$-$14–p.37, l.16; p.37, l.$-$15–l.$-$14; p.38, l.11–l.17; p.39, (1.42)–(1.46)], the concept of dipole layer is the key to understanding the physical meaning of Green's theorem or those of boundary conditions. This is the reason why Jackson discusses dipole layers [Jackson [44, §1.6]] before boundary conditions [Jackson [44, §1.8–§1.10]]. However, it is difficult to understand the former topic without knowing dipoles or point dipoles in advance. Therefore, it would be better prepared for understanding if one read Jackson [44, §1.6] again after being familiar with dipoles and point dipoles.
Remark. By Jackson [44, p.35, (1.31)] and Wangsness [86, p.36, (1-135)],
$|\nabla'^2(\frac{1}{|x-x'|}) = -4\pi\delta(x-x')$ [Jackson [44, p.36, l.$-$3]].

5.4 Physical ideas vs. their formal formulations

Physical ideas are usually simple, but their formal formulations in mathematical language can be sophisticated.

Example 5.2. (Watson's lemma)
Watson's lemma considers the integral $\int_0^\infty e^{-zt} f(t)dt$. The dominant value of the integral occurs near $t = 0$. This observation suggests that we estimate the integral by replacing f with its local expansion at $t = 0$. For the formal formulation of the lemma, see Koekoek [48, Theorem 2].

5.5 Physical proofs

5.5.1 A theorem's proof should be guided by its physical theme

Guided by a theorem's physical theme, one may develop a better strategy to prove it.

Example 5.3.
Coddington–Levinson [18, p.319, l.$-$15–l.$-$10] uses the following argument:
if [((A and C) $\Rightarrow B$) and ($B \Rightarrow C$)], then ($A \Rightarrow B$) ($*$), where
A = Coddington–Levinson [18, p.318, (1.16) & (1.17)];
C = ($|\varphi(t)| \leq \delta$ and Coddington–Levinson [18, p.319, (1.22)]) (see Coddington–Levinson [18, p.319, l.15–l.16]);
B = Coddington–Levinson [18, p.319, (1.23)].

If in $(*)$ we substitute C into B, we see that the conclusion $(A \Rightarrow C)$ is false. Thus Levinson's argument is incorrect. However, the hypothesis $[((A \text{ and } C) \Rightarrow B) \text{ and } (B \Rightarrow C)]$ ensures that if A holds, then B and C are equivalent. We can correct Levinson's mistake by the following method:

Even though the estimate provided by Pontryagin [65, p.211, l.−11] is poorer than that given in Coddington–Levinson [18, p.319, (1.23)], we may use the former estimate to prove C. Thereby, we can obtain the better estimate B.

Example 5.4. (The mathematical formulation of the second law of thermodynamics leads to a criterion for integrability of Pfaffian forms)[Sneddon [75, p.41, l.16–l.27; p.35, Theorem 8]; Zemansky–Dittman [91, p.169, (7-7); p.170, l.10–l.12; p.173, l.4–l.11]]

The algebraic criterion for integrability of Pfaffian forms [Sneddon [75, p.21, Theorem 5]] is good for calculation, while the geometric (or physical) criterion for integrability of Pfaffian forms [Sneddon [75, p.34, Theorem 7; p.35, Theorem 8]] is good for geometric (or physical) considerations. Pfaffian forms and thermodynamics are closely related. Without considering thermodynamics we cannot see the insight of Pfaffian forms; Without considering Pfaffian forms, we would have no mathematical foundation for thermodynamics. One should establish a solid connection between the two fields:

The connection from Pfaffian forms to thermodynamics: By Zemansky–Dittman [91, p.173, (7-13); p.174, (7-14)], the function μ is, apart from a multiplicative constant, a function only of the empirical temperature of the system [Sneddon [75, p.41, l.−6–l.−4]].

The connection from thermodynamics to Pfaffian forms: By Sneddon [75, p.41, l.−17–l.−16; l.−8], we see that the differential form for $đQ$ referring to a physical system of any number of independent coordinates possesses an integrating factor simply because of the second law of thermodynamics [Zemansky–Dittman [91, p.170, l.−14–l.−12]]. By Sneddon [75, p.19, Theorem 2], a system of two independent variables has a $đQ$ which always admits an integrating factor regardless of the second law [Zemansky–Dittman [91, p.173, l.12–l.13]].

The differences between thermodynamics and the general theory of Pfaffian forms: Zemansky–Dittman [91, p.173, l.8–l.10; l.14–l.16].

5.5.2 A physical proof is usually more direct than a geometric proof

Example 5.5. (The geometric criterion for integrability of the Pfaffian differential equation)

For Sneddon [75, p.35, Theorem 8], Carathéodory's thermodynamic proof is more direct than Born's geometric proof because the latter proof uses reduction to absurdity in Sneddon [75, p.38, l.−16–l.−9]. From the similarity between the path given in Sneddon [75, p.36, Fig. 11] and the integral path of Reif [68, p.160, (5·4·2)], we see that Carathéodory's idea originates from solving Reif [68, p.160, (5·4·1)]. His proof is closely related to the measurement of entropy using a quasi-static process. It would be difficult to understand the essence of Carathéodory's proof if one fail to know its physical meaning. In contrast, Born's proof involves only the geometric shape of solutions of the Pfaffian differential equation. One cannot use Born's method to measure entropies.

Remark. Continuously deforming the cylinder [Sneddon [75, p.38, l.−15–l.−14]] refers to reducing the cross section area of σ to 0. The band of accessible points [Sneddon [75, p.38, l.−13–l.−12]] refers to the segment IG_0.

5.5.3 Physical proofs are better than analytic proofs

Color painting adds more dimensions and varieties to black-and-white drawing. Similarly, physical and geometric proofs provide more meanings, pictures, insights, and interesting stories than analytic proofs.

Example 5.6. (Three proofs of the addition theorem for spherical harmonics: Watson–Whittaker [88, p.395, l.7–l.21], Cohen-Tannoudji–Diu–Laloë [19, vol. 1, pp.688–689], and Jackson [44, p.110, l.12–p.111, l.10])

In terms of the publishing dates of the above textbooks, the proofs of the later published books are better. The improvements are as follows:

(1) The choices of notations, coordinate systems, orthonormal functions become more compatible to the physical theme of the theorem.

  (a) Notations: For spherical harmonics, the notation given in Jackson [44, p.108, l.-11] is concise, while the notation given in Watson–Whittaker [88, p.392, l.-4] is awkward. The formula given in Jackson [44, p.110, (3.63)] is concise, while the formula given in Watson–Whittaker [88, p.393, l.-7] is awkward. The awkward formulas given in Watson–Whittaker [88, p.394, l.3–l.-9] may blur the essential ideas.

  (b) Orthonormal functions: Since we are discussing the solutions of Laplace's equation in spherical coordinates (Jackson [44, p.95, l.-13] and Watson–Whittaker [88, p.391, l.-1]), it is more appropriate to choose Y_{lm} on the unit sphere instead of P_n^m on $[-1,1]$ as the desired set of orthonormal functions (Jackson [44, p.108, l.16]).

(2) Ideally, the best physical proof is the one each of whose step has a pertinent physical interpretation. The development of physical methods shows the tendency toward such an ideal:

  (a) The choice of n given in Watson–Whittaker [88, p.395, l.12] lacks physical motivation, while the proof of Cohen-Tannoudji–Diu–Laloë [19, vol.1, p.688, (72)] supplies a physical reason: An eigenfuction of the angular momentum L^2 remains as an eigenfunction with the same eigenvalue after a rotation. The fact that the rotation operators commute with L^2 (Cohen-Tannoudji–Diu–Laloë [19, vol.1, p.688, l.-15–l.-14; p.699, (57)]) is more obvious than the fact that $\bigtriangledown^2$ is invariant under the rotation operators (Jackson [44, p.110, l.-8]).

  (b) Strictly speaking, Watson leaves a gap in the proof of the formula given in Watson–Whittaker [88, p.395, l.15]. Because θ_1' is a function of (θ,ϕ) and (θ',ϕ'), he should have expressed $P_n(\cos\theta_1')$ as an expansion of spherical harmonics in a form similar to that of the formula given in Cohen-Tannoudji–Diu–Laloë [19, vol.1, p.688, (74)]. If the expansion involved a term Y_{km}, where k is other than n, then he would not be able to derive the formula given in Watson–Whittaker [88, p.395, l.15]. Either poor notations or the lack of physical motivations fails him to detect the said gap.

  (c) Cohen-Tannoudji–Diu–Laloë [19, vol.1, p.689, (77)(i)] is derived from the fact that rotations form a group. Cohen-Tannoudji–Diu–Laloë [19, vol.1, p.689, (79)] is the Schwartz inequality. Therefore, the discussion given in Cohen-Tannoudji–Diu–Laloë [19, vol.1, pp.688–689, §γ (iii)] is purely analytical. Actually, its idea is worse than that given in Watson–Whittaker [88, p.395, l.12–l.17]. Consequently, Jackson [44, §3.6] replaces it with Jackson [44, p.111, l.1–l.6] which reveals more insights about rotation and angular momentum.

  (d) If we correct the above shortcomings and make the following changes, the proof given in Jackson [44, §3.6] would be perfect.

$[4\pi(2l+1)^{-1}]^{1/2}Y^*_{lm}(\theta(\gamma,\beta),\phi(\gamma,\beta)) = \Sigma^l_{m=-l}A_{lm}Y_{lm}(\gamma,\beta)$ (Jackson [44, p.109, (3.58)] and Cohen-Tannoudji–Diu–Laloë [19, vol.1, p.688, l.-12]).
Let $\gamma = 0$. We have $[4\pi(2l+1)^{-1}]^{1/2}Y^*_{lm}(\theta',\phi') = A_{l0}$
$= A_m(\theta',\phi')$ (Jackson [44, p.366, (3.66) and (3.60)]).

5.6 A proper physics model can be a natural guide to the study of PDEs

In order to complete the study of wave equations, we must consider the following three cases. For each case, we have to choose a proper physics model as a guide.
Case 1. 1 dimensional case, rectangular coordinates: continuous strings [Marion [55, §13.4–§13.8]].
Case 2. 2 or 3 dimensional case, polar or cyclindrical coordinates: circular membranes [Asmar [5, §4.2–§4.3]].
Case 3. 3 dimensional case, spherical coordinates: electric potentials [Jackson [44, §3.1–§3.6]].

Each case lays the basis for studying the next one. The proper model is a template for all other similar models. Physics models and solutions of PDEs are inseparable and complement each other. Without a physics model as a guide, PDEs become dull and abstract. Only through a model may we propose significant questions and effectively find their solutions. Without choosing significant boundary and initial conditions, the boundary value problems can become practically meaningless. Although electric potentials can also be used as a physics model for Case 2 [Jackson [44, §3.7–§3.8]], they are not as good as circular membranes because we cannot see the former.

6 Improvements of classical methods

If there are drawbacks in a classical method, all we have to do is provide ideas to improve them. If there is a gap in its proof, we simply fill the gap. In other words, a remedy rather than a thorough revamp is all we need. This introduction mode based on needs may make the key to improvement most outstanding.

Example 6.1. (Precision improvement of a classical method)

Let S be a ruled surface [Bell [6, p.313, l.-1]]. If $\alpha'b' - \beta'a' = 0$ [Bell [6, p.314, l.15; p.313, l.-5]], then S is developable [Bell [6, p.314, l.15–l.16]].

Proof. We have $d = O(\delta t^3)$ in Bell [6, p.314, l.12–l.13], but Bell jumps to the conclusion that $d = 0$. Thus, there is a gap needed to be filled.
Let the directrix of S be $y(s) = (\alpha(s),\beta(s),0)$, $z(s) = (a(s),b(s),1)$, and
$x(s,t) = y(s) + tz(s)$.
S is developable
$\Leftrightarrow 0 = |y'zz'| = \begin{vmatrix} \alpha' & \beta' & 0 \\ a & b & 1 \\ a' & b' & 0 \end{vmatrix} = \beta'a' - \alpha'b'$ [Kreyszig [49, p.169, Theorem 59.1]]. □

Remark. In the above proof, we use the method of differential geometry to fill the gap of a classical proof. Thus, we see the advantage of modern geometry. At the same time, we also see the concept of "consecutive" generators is useful to the intuitive understanding of a ruled surface although it is difficult to make its definition rigorous. Consequently, classical geometry and differential geometry are complementary to each other.

In analytic geometry, we discuss geometry with coordinate systems. Geometry is our main study goal and coordinate systems are nothing but tools to express geometric objects as equations. We should choose the coordinate system that makes the equation of the main geometric object in the simplest form. This approach will allow us to reduce calculations, to easily recognize its properties, etc. For example, when discussing plane sections of a conicoid, we should express these conics in standard form intead of general form.

Example 6.2. (Standard form vs. general form)
All parallel plane sections of a conicoid are similar and similarly situated conics [Bell [6, p.74, Ex. 3]].

Proof 1. Bell [6, p.74, l.12–l.16]
By Fine–Thompson [28, p.137, l.-3–l.-1], the centers of resulting conics are collinear and vary with a, h, g, b, f.
By Fine–Thompson [28, p.137, l.11], the axes of a resulting conic is determined by λ, which is, in turn, determined by a, b, h. Thus, the axes of every resulting conic make the same angles with plane coordinate axes, so the conics are similarly situated.
By Fine–Thompson [28, p.138, l.-17], the conics are similar. □

Proof 2. Bell [6, p.134, l.12–p.135, l.19] □

Remark. In the first proof, we express a plane in simple form $z = k$; in the second proof, we express a plane in general form $lx + my + nz = p$. In the first proof, our goal is to find the standard form of a conic, and then its axes and direction-cosines of the axes. The goal leads directly to solutions; the approach helps us see the insight and key ideas. The second proof relies on the comparison between Bell [6, §86, (2) & (3)] and Bell [6, §87, (2) & (3)]. These formulas are derived from the necessary and sufficient conditions for a plane to touch a cone given in Bell [6, p.120, l.1–l.4]. Consequently, the second proof is not as simple and direct as the first one. This example shows that simplification and standardization are the keys to effective studying analytic geometry.

Example 6.3. (Simplifying complicity: one right coordinate system does it all)
(Intersection of three planes) Bell [6, §45]
Consider the system of equations given in Bell [6, p.49, (1), (2), & (3)]. Let
$r =$ rank of the coefficient matrix $\begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix}$ and
$r' =$ rank of the argumented matrix $\begin{pmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \end{pmatrix}$.

Table 1: The intersection of three planes

Systems	Case number	Algebraic Classification	Geometric classification
Consistent	1	$r = 3$	Three planes intersect at one point.
	2	$r = r' = 2$; no two rows of the argumented matrix are proportional.	Three planes intersect in one line.
	3	$r = r' = 2$; two rows of the argumented matrix are proportional.	Two planes are coincident, and the third cuts the others.
	4	$r = r' = 1$	All three planes are coincident.
Inconsistent	5	$r = 2, r' = 3$; no two rows of the coefficient matrix are proportional.	Normals are coplanar, planes intersect in pairs, and the intersecting lines form a triangular prism.
	6	$r = 2, r' = 3$; two rows of the coefficient matrix are proportional, but the same two rows of the argumented matrix are not proportional.	Two parallel planes intersect a third plane.
	7	$r = 1, r' = 2$; no two rows of the argumented matrix are proportional.	All planes are parallel and distinct.
	8	$r = 1, r' = 2$; two rows of the argumented matrix are proportional.	Two planes are coincident, and the third is parallel.

Proof. Based on geometric considerations, there are no cases other than the above eight cases. In order to prove that the two corresponding classifications are equivalent, all we have to do is find the coordinate system to put a case in simplest equation form and then determine the ranks. Because ranks are invariant under translations and rotations and the general case can be obtained from a simple case by a finite number of translations and rotations, it is unnecessary to consider the general case. For example, for case 5, all we have to do is consider $y = a, y = \alpha x, y = \beta x$, where $a \neq 0$ and $\alpha \neq \beta$. □

Remark. Bell [6, §45] attempts to prove the same thing, but it chooses the hard way. In this context, the emphasis should be on geometry rather than matrix theory. Bell might know some matrix theory, but he failed to master it or make good use of it.

Example 6.4.

Fine–Thompson [28, pp.60–61, §79 A–C] give three derivations of equation of tangent. In fact, they are all derived from the viewpoint of calculus: the tangent line is the limit of secant lines. Thus, the three derivations are just three ways of constructing a secant line of a conic. From the viewpoint of differential geometry, in order to find the tangent line, we would find the normal first [O'Neill [59, p.127, Theorem 1.4; p.148, Lemma 3.8]] because in three dimensions the normal determines the tangent plane. The differential-geometric approach is more direct.
Remark. The same discussion applies to Fine–Thompson [28, pp.81–83, §102 A–C].

Example 6.5. (Corresponding versions of the same idea motivate us to find the general case)

(a) Polars versus polar planes

 (i) 2-dim: The polar of a point with respect to a conic [Fine–Thompson [28, p.148, l.16]]
 (ii) 3-dim: The polar plane of a point with respect to a conicoid [Bell [6, p.104, l.−6–l.−5]]

(b) Symmetry between two poles [resp. polar lines]

 (i) 2-dim: The polar of P_1 passes through P_2 [Fine–Thompson [28, p.148, l.−7–l.−6]]
 (ii) 3-dim: The polar plane of (α, β, γ) passes through (ξ, η, ζ) [Bell [6, p.105, l.−12–l.−10]]; the polar plane of any point on a line AB passes through a line PQ [Bell [6, p.105, l.−10–l.−5]]

The general case: all the above concepts or statements can be generalized to n-dimensional manifolds.

Example 6.6. (Intergration of algebra and geometry)

Fine–Thompson [28, §274] provides an algebraic derivation of the equation for the plane through three given points. It seems that Fine talks shop all the time. Actually, we may also give a geometric derivation. The direction-cosines of the plane's normal are proportional to
$(x_2 - x_1, y_2 - y_1, z_2 - z_1) \times (x_3 - x_1, y_3 - y_1, z_3 - z_1)$.
Consequently, the equation for the plane is
$(x_2 - x_1, y_2 - y_1, z_2 - z_1) \times (x_3 - x_1, y_3 - y_1, z_3 - z_1) \cdot (x - x_1, y - y_1, z - z_1) = 0$. Namely,
$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = 0$ [Kreyszig [49, p.17, (5.14)]].

Remark. Note that $\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = \begin{vmatrix} x - x_1 & y - y_1 & z - z_1 & 0 \\ x_1 & y_1 & z_1 & 1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 & 0 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 & 0 \end{vmatrix} = \begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix}$.
These equalities integrate the algebraic and geometric meanings of the equation for the plane and explain the reason why the proof in Fine–Thompson [28, §276A] is equivalent to that in Fine–Thompson [28, §276C]. The above geometric derivation makes the proof of the statement given in Fine–Thompson [28, p.209, l.11–l.12] become easy.

Example 6.7. (Cartesian coordinates lack ability to distinguish infinities of all directions)

The center is (∞, ∞, ∞) [Fine–Thompson [28, p.285, l.−7]].

Proof. The algebraic proof follows from Fine–Thompson [28, p.266, (8)]. A geometric proof proceeds as follows: Let the center of conicoid be (x_0, y_0, z_0). Since (x_0, y_0, z_0) is the common midpoint of chords of the parabola on the plane $y = 0$ [Fine–Thompson [28, p.243, Figure]], $y = 0, x_0 = \infty = z_0$. The line through $(\infty, 0, \infty)$ with direction cosines $(\lambda, 0, \nu)$ is $\frac{x-\infty}{\lambda} = \frac{y-0}{0} = \frac{z-\infty}{\nu}$, which can also be interpreted as
$\frac{x-\infty}{\lambda} = \frac{y-\infty}{0} = \frac{z-\infty}{\nu}$, the line through (∞, ∞, ∞) with direction cosines $(\lambda, 0, \nu)$. This is why we find from the algebraic proof that the distance between (∞, ∞, ∞) and any point on the paraboloid is the same. Consequently, from the algebraic viewpoint, $(\infty, 0, \infty) = (\infty, \infty, \infty)$. Since (x_0, y_0, z_0) is the common midpoint of chords of the parabola on the plane $x = 0$ [Fine–Thompson [28, p.243, Figure]], $x = 0, y_0 = \infty = z_0$. □

Remark. The paraboloid has center at (∞, ∞, ∞) because we use an improper tool. Cartesian coordinates lack ability to distinguish infinities of all directions. If we use spherical coordinates instead, then $(\theta, \phi, r = \infty)'s$ will represent different points if $(\theta, \phi)'s$ point to different directions. In this case, there will be no common midpoint for the chords through the origin.

Example 6.8. (If the set of centers is empty and we allow a point involving ∞ to be its element due to a tool abuse, then all the theorems to which the false existence of elements leads will be meaningless [Bell [6, §152–§153]])

Consider the system of equations given in Bell [6, p.216, (1), (2), & (3)]. Let

$r =$ rank of the coefficient matrix $\begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix}$ and

$r' =$ rank of the argumented matrix $\begin{pmatrix} a & h & g & u \\ h & b & f & v \\ g & f & c & w \end{pmatrix}$.

The set of centers may be any of the following cases:

(i) A point: $r = 3$ [Table 1], (ellipsoid, hyperboloid, or cone).

(ii) A line: $r = r' = 2$; no two rows of the argumented matrix are proportional (elliptic or hyperbolic, cylinder, pair of intersecting planes).

(iii) A plane: $r = r' = 1$ (pair of parallel planes).

(iv) The empty set $\varnothing$.

The classification of cases for centers should be determined by the final solution of a system of equations rather than the solution process. Consider the parabola $x^2 = 4ay$ with polar coordinates. Let $A = (\cos\pi/4, \sin\pi/4), B = (\cos\pi/3, \sin\pi/3)$. The middle point for the chord along $\overrightarrow{OA} = (\infty, \pi/4) \neq (\infty, \pi/3) =$ the middle point for the chord along $\overrightarrow{OB}$. We may find the algebraic solution (∞, ∞) for the center [Example 6.7] because the Cartesian coordinates are inadquate to tell the above difference. In this example, the equation for the second central plane is $4a = 0$, which is the empty set instead of a plane. Thus, this standard type becomes an exception for the classification given in Bell [6, p.216, l.-2–p.217, l.16]. If the set of centers is empty and we allow a point involving ∞ to be its element due to a tool abuse, then all the theorems to which the false existence of elements leads will be meaningless.

On the one hand, a general theory without giving examples lacks concrete pictures. On the other hand, if we focus on examples alone or the discription of examples fails to be consistent with the direction of the general theory's development, then we may not be able to clearly see the direction of the theory's development.

Example 6.9. (Examples vs. general theory: Coordinate systems)

A system of unit vectors $\hat{\rho}, \hat{\varphi}, \hat{z}$ [Symon [79, p.95, l.-1; Fig. 3.22]] gives the orthonormal basis of cylindrical polar coordinates; a system of unit vectors $\hat{r}, \hat{\theta}, \hat{\varphi}$ [Wangsness [86, p.31, l.-11–l.-10; p.32, Figure 1-39]] gives the orthonormal basis of spherical coordinates. For the general case, see Arfken & Weber [2, p.8, Exercise 2.1.1].

Example 6.10. (Using the standard form of conicoids to simplify the proofs of theorems about symmetric matrices)

Consider the system of equations given in Bell [6, p.216, (1), (2), & (3)]. Let

$r =$ rank of the coefficient matrix $\begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix}$ and

$r' =$ rank of the argumented matrix $\begin{pmatrix} a & h & g & u \\ h & b & f & v \\ g & f & c & w \end{pmatrix}$. Let

$D = \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix}$ and $\Delta = \begin{vmatrix} a & h & g & u \\ h & b & f & v \\ g & f & c & w \\ u & v & w & d \end{vmatrix}$.

Bell [6, p.220, l.21–l.22] says that $[(r = 2, r' = 3) \Rightarrow$ (the conicoid is a paraboloid)]. Fine–Thompson [28, p.283, Table] says that $[(D = 0, \Delta \neq 0) \Rightarrow$ (the conicoid is a paraboloid)]. Hence $(r = 2, r' = 3) \Leftrightarrow (D = 0, \Delta \neq 0)$.

Proof. Since the rank and the determinant of a matrix are invariant under nonsingular linear transformations, we may use standard forms of conicoids to prove this theorem. Using a complete list [Fine–Thompson [28, §237]] of standard forms of conicoids to check if there is any form satisfies the property $(r = 2, r' = 3)$, we find that only the standard form $\frac{x^2}{a^2} \pm \frac{y^2}{b^2} = \frac{2z}{c}$ of a paraboloid satisfies this property. Consequently, paraboloids can be characterized by the property $(r = 2, r' = 3)$. □

Example 6.11. (Trinity of consecutive points, approach to the same point, and contact of higher order)

The definition given in Bell [6, p.279, l.−12–l.−10] is a rigorous statement of the definition given in Weatherburn [90, vol.1, p.12, l.6–l.8]. The definitions of osculating plane, osculating circle, osculating sphere given in Bell [6, p.279, l.−12–l.−10; p.292, l.1–l.3; l.−3–l.−1] are based on the same idea of osculation. The second approach is heuristic, systematic and unified. In contrast, the definitions given in Kreyszig [49, p.33, Table 10.1; p.51, l.4 & l.17] look artificial. The common idea of these three definitions becomes vague in the third approach. The important step for the construction process (the circle *PQR* [Bell [6, p.292, l.2]] or the sphere *PQRS* [Bell [6, p.292, l.−2]]) in the second approach is lost in the third approach and cannot be restored by using the results in the third approach alone. From hindsight, consecutive points can be considered a brief expression for contact of higher order [Kreyszig [49, p.50, l.−6; p.51, l.6]]). The contact of second or third order can easily be generalized to the nth order.

Example 6.12. (Regarding a rigorous proof as an improvement)

Sneddon [75, p.20, l.1–p.21, l.7] provides a rigorous proof of Bell [6, p.318, l.−9–l.−1].

Remark. When reading classical books, one should be familar with their frequent mistakes and should know how to correct them. Otherwise, one cannot appreciate these books. In most cases, the statement of a theorem is correct, but the author fails to provide a rigorous proof. If we read Bell [6, p.318, l.−9–l.−1] alone, we do not know the proof can be improved. Likewise, if we read Sneddon [75, p.20, l.1–p.21, l.7] alone, we do not know it is an improvement of a classical theorem.

Example 6.13. (Only after a construction is tailored to our needs can it solve the problem effectively)

In order to prove the countable additivity of P, the construction given in the proof of Borovkov [13, p.31, Theorem 1] is tailored to our needs and effectively meets our goal. The construction is lean, simple and clear. Note that we can take $\tilde{B}_n$ such that $B_{n+1} \subset \tilde{B}_n \subset B_n$. In contrast, the step IV of the proof of Rudin [72, p.42, Theorem 2.14] provides a second construction, but this construction is too general and abstract to be practical. The construction is burdened by unnecessary equipment: the existences in the proofs of Rudin [72, p.37, Theorem 2.5; p.38, Theorem 2.7] are provided by topological axioms; the construction in the proof of Rudin [72, p.40, Lemma p2.12] is too complicated to be useful in practice. Consequently, it is difficult to apply the second construction to practical cases. A big apparatus may be impressive from the theoretical viewpoint. However, it not only fails to point out the key idea, but also is useless in applications.

Example 6.14. (Only after a proof is tailored to our needs may we grasp the key idea)

The more direct a proof is, the more powerful it is. The less deviated or involved an argument is, the clearer the key idea becomes. There are three proofs of Borovkov [13, p.435, Lemma 3]. Borovkov [13, p.436, l.1–l.10] provides the first proof. The second one follows from Borovkov [13, p.115, Lemma 2]. The third one follows from Rudin [72, p.176, Theorem 8.17; Theorem 8.16; p.132, Theorem 6.11]. The first proof is the most direct of the three.

Example 6.15. (An index set must be chosen properly: one more candidate would be too many and one less would be too few)

The ch.f. of a random variable uniquely determines its distribution function [Borovkov [13, p.139, l.18–l.19]].

Proof. Let $n=2$ and $\Delta=(a_1,b_1)\times(a_2,b_2)$ [Borovkov [13, p.130, l.6–l.7]].
In order to define $F_\xi(0)$, we need to find Δ_i so that $\cup_i\Delta_i=(-\infty,0)\times(-\infty,0)$, where Δ_i's are mutually disjoint. This will ensure that except for countable i, $P_\xi(\partial\Delta_i)=0$. Then, by left-continuity of F_ξ and the inversion formula [Borovkov [13, p.130, l.6–l.9]], we may define $F_\xi(0)=\sup_i P(\xi\in\Delta_i)$.
Case I. (Improper choice: one more candidate would be too many)
If we define $F_\xi(0)=\sup_{\Delta\subset(-\infty,0)\times(-\infty,0),P_\xi(\partial\Delta)=0}P(\xi\in\Delta)$, at best we define only a unsolved problem. This is because many recruited candidates are unqualified, but we put them into consideraton and have no way to rid them in order to satisfy the condition. Confucius says, "Only by careful distriguishing what one knows from what one dones't may one have a deeper understanding."
Case II. (Wrong choice: one less would be too few)
$\forall k\in\mathbb{N}$, let $\Delta_k=(-k,-\frac{1}{k})\times(-k,-\frac{1}{k})$. Suppose we define $F_\xi(0)=\sup_k P(\xi\in\Delta_k)$. Since $P_\xi(\partial\Delta_k)$ may not be 0, there may be not enough Δ_k's that can satisfy the condition $\cup_{P_\xi(\partial\Delta_k)=0}\Delta_k=(-\infty,0)\times(-\infty,0)$. Thus, we choose too few candidates.
Case III. (Proper choice)
$\forall x\geq 1$, let $\Delta_x=(-x,-\frac{1}{x})\times(-x,-\frac{1}{x})$. By Rudin [72, p.17, Theorem 1.19(d)], we define $F_\xi(0)=\sup_{x\geq 1,P_\xi(\partial\Delta_x)=0}P(\xi\in\Delta_x)$. Then Δ_x's are mutually disjoint. Consequently, except for countable x, $P_\xi(\partial\Delta_x)=0$. $\square$

Remark. Mathematics discusses the process of finding a solution rather than just proves that a given solution is true.

Example 6.16. (The strong law of large numbers and the central limit theorem [Borovkov [13, p.151, Theorem 1; p.152, Theorem 2]; Lindgren [53, p.155, Khintchine's theorem; p.158, central limit theorem]])

(1) The proofs of both Lindgren [53, p.155, Khintchine's theorem; p.158, central limit theorem] and Borovkov [13, p.151, Theorem 1; p.152, Theorem 2] are essentially the same except that the former proofs use the lemma given in Lindgren [53, p.156, l.6–l.8] while the latter proofs do not. It is easy for the former proofs to be generalized to the multidimensional case, but it is difficult for the latter proofs. Furthermore, $-\frac{t^2}{2}+o(1)\to-\frac{t^2}{2}$ given in Borovkov [13, p.153, l.3] is incorrect because $o(1)$ refers to $t\to 0$ rather than $n\to\infty$.

(2) (Stronger convergences) The proofs of both Lindgren [53, p.155, Khintchine's theorem; p.158, central limit theorem] and Borovkov [13, p.151, Theorem 1; p.152, Theorem 2] use Borovkov [13, p.132, Theorem 2]. Therefore, the covergences in both Khintchin's theorem and the central limit theorem are essentially weak convergences. The proof of Borovkov [13, p.151, Theorem 1] gives the weak covergence $F_{S_n/n}\Rightarrow a$. By Lindgren [53, p.154, Theorem B], the weak convergence can be strengthened to

the convergence in probability $S_n/n \xrightarrow{p} a$. The strong law of large numbers [Chung [13, p.133, Theorem 5.4.2 (8)]] strengthens the convergence in probability $S_n/n \xrightarrow{p} a$ further to the almost sure conergence $S_n/n \xrightarrow{a.s.} a$ [Borovkov [13, p.151, l.-11]]. The proof of Borovkov [13, p.152, Theorem 2] gives the weak covergence $F_{\zeta_n} \Rightarrow \Phi$. By Borovkov [13, p.116, Theorem 6], we have the pointwise convergence $F_{\zeta_n}(x) \to \Phi(x)(x \in \mathbb{R})$. By Parzen [60, p.438, Exercise 5.2], $F_{\zeta_n}(x) \to \Phi(x)$ uniformly in $x \in \mathbb{R}$.
Remark. The strong law of large numbers for Bernoulli scheme follows from Borovkov [13, p.91, Theorem 2; p.109, Theorem 1].

(3) If the metric space S given in Billingsley [9, p.3, l.9] is the real line $\mathbb{R}$, then the proof given in Lindgren [53, p.154, l.-5–p.155, l.5] is more intuitive than that given in Billingsley [9, p.27, l.5–l.15].
Remark. For the right side of the inequality given in Lindgren [53, p.155, l.2], note that $P(Y_n = k - \varepsilon) = 0$ [Billingsley [9, p.26, Theorem 2.1(iii)]].

(4) The motivation to choose Z_n in formulating the central limit theorem [Lindgren [53, p.157, l.8–l.-3]].
I. Choose Z_n instead of S_n to keep track of the shape of the limiting distribution function.
II. Choose Z_n instead of Y_n to avoid the singularity of the limiting distribution function.
By the weak law of large numbers, $F_{Y_n} \Rightarrow F_{I_{EX}}$, where $F_{I_{EX}}$ has a single jump at EX.
III. Standization (mean = 0, var = 1) that keeps the limiting distribution function from shrinking or expanding leads us from Y_n to Z_n naturally.

(5) (Motivation for using characteristic functions; key points vs. details; natural proofs; physical meanings)
Both Reif [68, p.35, l.6–p.40, l.8] and Borovkov [13, p.75, Theorem 7; §5.1–§5.3; §5.5; §7.1–§7.4; §7.6; §8.1–§8.2] discuss the strong law of large numbers and the central limit theorem. The former indicates the motivation for using characteristic functions to prove these theorems [Reif [68, p.36, l.1–l.17]] and reveals that the key idea of proving these theorems is simple, original and excellent [Reif [68, p.36, l.17–l.-1]]. However, the former lacks details; its statements are crude; its proofs are not rigorous. Although the latter provides details, accuracy, and rigor, its proofs lack motivations and its key points are vague. The way to keep the merits of both approaches is to select the key statements in the former and find their corresponding rigorous ones in the latter.
I. $P(x)dx$ [Reif [68, p.35, l.10]] $\to dF(x)$ [Borovkov [13, p.29, l.-17]]. For the latter expression, every component (the set of elementary outcomes, σ-algebra, probability) of the probability space [Borovkov [13, p.17, Definition 6]] is clearly specified. Thus, we have a rigorous mathematical structure ready to hand.
II. Reif [68, p.35, (1·10·2)] $\to$ Borovkov [13, p.54, l.7–l.16; p.126, l.11]. This can lead to the equality given in Borovkov [13, p.97, l.8].
III. By using Dirac δ function, we obtain Reif [68, p.35, (1·10·2) & (1·10·3)] = Reif [68, p.36, (1·10·4)]. Reif [68, p.36, (1·10·5)] $\to$ Borovkov [13, p.126, l.4–l.6; p.130, (5)]. Reif [68, p.36, (1·10·5)] which leads to Reif [68, p.36, (1·10·6)] gives the reason why we should use characteristic functions to prove the strong law of large numbers and the central limit theorem. Consequently, do not consider the inversion formula a unnatural thing. In fact, based on the physical consideration given in Reif [68, p.36, l.1–l.17], only through the use of characteristic functions and the inversion formula may we have a simple, natural and general [Reif [68, p.37, l.-7–l.-6]] method of dealing with the convergence of the sum of a sequence of independent identically distributed random variables. The approaches given in Borovkov [13, p.97, l.-12–p.99, l.11] and in the proof of Chung [17, p.114, Theorem 5.2.2] are artificial, while the proofs of Borovkov [13, p.151, Theorem 1; p.152, Theorem 2] are natural.
IV. For the Riemann–Lebesgue lemma [Borovkov [13, p.129, 8]; Rudin [72, p.197, Theorem 9.6)]], Reif [68, p.38, l.1–l.4] provides its physical meaning and the motivation for its formulation. The proof given

in Reif [68, p.38, Remark] is not as good as the proof given in `https://en.wikipedia.org/wiki/Riemann%E2%80%93Lebesgue_lemma`.
V. Reif [68, p.38, l.1–l.12] provides the motivation for formulating the central limit theorem [Reif [68, p.39, l.-7]; Borovkov [13, p.152, Theorem 2]].

Example 6.17. (Indigo blue is extracted from the indigo plant but is bluer than the plant it comes from)

Most mathematical theorems do not come from nowhere. A new theorem is often a supplement, a stronger version, an analog or an extention of an old theorem. The generation of this kind of derivatives makes up a significant part of the development of a theory. Thus, indigo blue is extracted from the indigo plant but is bluer than the plant it comes from. The following evidences convince us that we should learn to control rather than follow the flow of a proof.
I. Supplements: Chung [17, p.133, Theorem 5.4.2, (9)] is a supplement of Chung [17, p.133, Theorem 5.4.2, (8)]. The former discusses Case $\mathscr{E}(|X_1|) = \infty$, while the latter discusses Case $\mathscr{E}(|X_1|) < \infty$.
II. Stronger versions: Chung [17, p.133, Theorem 5.4.2] is stronger than Chung [17, p.114, Theorem 5.2.2]. The convergence in probability of the latter is strengthened to the almost sure convergence of the former.
III. Extensions: Chung [17, p.134, Theorem 5.4.3] is an extension of Chung [17, p.133, Theorem 5.4.2, (9)] because the hyperthesis of the former is more flexible than that of the latter. Similarly, Chung [17, p.121, Theorem 5.3.1] is an extension of Chung [17, p.50, Chebyshev's inequality] and Chung [17, p.116, Theorem 5.2.3] is an extension of Chung [17, p.114, Theorem 5.2.2]. The example of strengthening given in II is a special case of Chung [17, p.126, Theorem 5.3.4]. Since Borovkov [13, p.151, Theorem 1] is the same as Chung [17, p.114, Theorem 5.2.2], we expect the proof of Borovkov [13, p.151, Theorem 1] and that of Chung [17, p.133, Theorem 5.4.2, (8)] should be similar, but they are actually different. In order to organize the structure of the proof of Chung [17, p.133, Theorem 5.4.2, (8)], we may use Borovkov [13, p.151, Theorem 1] and Chung [17, p.126, Theorem 5.3.4] to prove Chung [17, p.133, Theorem 5.4.2, (8)]. The proof provided by this method is more compatible with that of Borovkov [13, p.151, Theorem 1] than that of Chung [17, p.133, Theorem 5.4.2, (8)].
IV. Analogs: The formula given in Ellison–Ellison [27, p.265, l.-10] is an analog of Ellison–Ellison [27, p.46, Theorem 2.5].
Remark 1. The proof of Chung [17, p.133, Theorem 5.4.2, (8)] and that of Loéve [54, p.251, l.15–p.252, l.6] are essentially the same.
Remark 2. (Control the work flow by dividing it into several applicable sections) It seems magical that Chung [17, p.121, Theorem 5.3.1] makes the hypothesis of Borovkov [13, p.75, (13)] more flexible [Chung [17, p.121, l.-3–l.-1]]. If we compare Chung [17, p.122, l.-2–l.-1] with Borovkov [13, p.75, l.4], we find that their key ideas are essentially the same except that the former divides the work flow into several applicable sections [Chung [17, p.122, l.10]]. This technique of segmentation designed for independence is also used in the proofs of Chung [17, p.123, Theorem 5.3.2; p.126, Theorem 5.3.4].

Example 6.18. (An advantaneous viewpoint can facilitate the calculations in a proof)

The proof [Pontryagin [65, p.50, l.-22–p.52, l.-4]] of the first part of Pontryagin [65, p.52, (B)] originates from the viewpoint of differentiation [Pontryagin [65, p.50, l.-14]]. In contrast, the proof [Hartman [40, p.324, l.1–l.16]] of Hartman [40, p.324, (1.15)(i)] originates from the viewpoint of changing variables [Hartman [40, p.324, l.11]]. The latter viewpoint can facilitate the calculations in the proof of Pontryagin [65, pp.50–51, Theorem 5].

Example 6.19. (Method vs. calculation for solutions)

$u(t)$ and $v(t)$ are linearly independent solutions of (2.1) if and only if $c \neq 0$ in (2.7) [Hartman [40, p.327, l.8–l.9]]

Proof with theory as a guide. $c \neq 0$
$\Leftrightarrow \det X(t) \neq 0$ [Hartman [40, p.326, (2.7)]]
$\Leftrightarrow ((u(t), p(t)u'(t))$ and $(v(t), p(t)v'(t))$ are linearly independent)
$\Leftrightarrow (u(t)$ and $v(t)$ are linearly independent) [Hartman [40, p.326, (iv)]].

□

Proof with calculations in mind. See Ince [43, p.116, l.14–p.118, l.16]. □

Remark. The first proof helps us grasp the key points, while the second proof helps us understand the original approach. The second proof requires patience and tricks [Hartman [40, p.326, l.12–l.13]] to deal with nuisances [Ince [43, p.117, l.14]; signs of cofactors] and difficult points [Ince [43, p.117, l.−20]; continuity].

Example 6.20. (Boltzmann's entropy formula)

Most textbooks in statistical mechanics define entropy artifically as $S = k \ln \Omega$. Actually, it is more heuristic to ask how to derive this formula. Consider the isothermal expansion of ideal gas.
$\Delta E = 0$ [Reif [68, p.126, (3·12·11)]]
$\Rightarrow \Delta Q = \Delta W = \int_{V_1}^{V_2} \bar{p} dV$
$= NkT \ln \frac{V_2}{V_1}$ [Reif [68, p.125, (3·12·8)]].
$dS = \frac{\Delta Q}{T} = k \ln \frac{\Omega(E,V_2)}{\Omega(E,V_1)}$ [Reif [68, p.64, (2·5·14)]].
It is more natural to relate entropy in thermodynamics to microstates at this moment and in this way.

Example 6.21. (An easy way to make the discussion of $\delta(x)$ rigorous)
$\delta(\phi - \phi') = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{im(\phi-\phi')}$ on $[-\pi, \pi)$ [Jackson [44, p.125, l.13, (3.139)(ii)]]

1st proof. $D_n(x) = \frac{\sin(n+1/2)x}{\sin(x/2)}$ [Rudin [70, p.174, (77)]]
$\Rightarrow D_n(2x) = \frac{\sin(2n+1)x}{\sin x}$
$\Rightarrow \frac{1}{\pi} \lim_{n\to\infty} D_n(2x) = \delta(x)$ [Born–Wolf [12, p.897, (20) & (21)]]
$\Rightarrow \frac{1}{\pi} \sum_{n=-\infty}^{\infty} e^{in(2x)} = \delta(x)$ [Rudin [70, p.174, (76)(i)]]
$\Rightarrow \frac{1}{\pi} \sum_{n=-\infty}^{\infty} e^{inx} = \delta(x/2) = 2\delta(x)$ [Cohen-Tannoudji–Diu–Laloë [19, vol. 2, p.1471, (20)]]. □

2nd proof. Let f be a continuous function of bounded variation with a period 2π.
$s_n(f;x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-t) D_n(t) dt$ [Rudin [70, p.175, (82)]]
$\Rightarrow \int_{-\pi}^{\pi} f(x-t) (\frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{imt}) dt = \lim_{n\to\infty} s_n(f;x) = f$ [Royden [69, p.232, Proposition 18]; Zygmund [92, vol. 1, p.57, Theorem (8.1)(ii)]]
$= \int_{-\pi}^{\pi} f(x-t) \sum_{q=-\infty}^{\infty} \delta(t - 2\pi q) dt$ [Cohen-Tannoudji–Diu–Laloë [19, vol. 2, p.1473, (31)]]
$= \int_{-\infty}^{\infty} f(x-t) \delta(t) dt$
$= \int_{-\pi}^{\pi} f(x-t) \delta(t) dt$ [$\delta(t) = 0$ if $t \notin [-\pi, \pi)$].
$\Rightarrow \delta(t) = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} e^{imt}$ a.e. on $[-\pi, \pi)$ [Rudin [72, p.31, Theorem 1.39(b)]] □

Remark 1. Strictly speaking, the first proof is not good because it fails to consider the requirement given in Reif [68, p.614, l.−7]. Both proofs are not rigorous because a generalized function should not be treated as a function. In order to correct the problem, we should use the language of functional analysis. Actually, the required supplement is not much. For the discussion of the Dirac delta function, it requires only Rudin [71, p.142, l.−6, (1); p.155, (2) & (5)] to bridge the gap between a function and a generalized function.

For the discussion of derivatives of δ [Cohen-Tannoudji–Diu–Laloë [19, vol. 2, p.1476, b]], it requires only Rudin [71, p.144, (1), (2) & (3)] to bridge the gap between a function and a generalized function.
Remark 2. The discussion given in Cohen-Tannoudji–Diu–Laloë [19, vol. 2, Appendix II] is not good because a generalized function should not be treated as a function. It requires a rigorous theory to correct and support the discussion. The theory contained in Rudin [71, chap. 6] is rigorous, but it fails to directly apply to the Dirac delta generalized function. Many physicists fail to understand the theory. This is the reason why theory and applications are easily disconnected. Therefore, it is important to identify their connections.

Example 6.22. (A good theorem should provide complete information)

(A) If we treat $\mathbb{R}^2$ as a topological subspace of its one-point compactification S^2 and denote the boundary relative to S^2 as ∂_∞, then the geometric meaning of "$y(t)$ tends to ∂E as $t \to \omega+$" given in Hartman [40, p.13, l.5–l.7] is "$(t, y(t))$ tends to $\partial_\infty E$ as $t \to \omega+$". A good theorem should provide complete information. In the above sense, the conclusion of Hartman [40, pp.12–13, Theorem 3.1] gives a complete geometric picture, while the result given in Hirsch–Smale–Devaney [23, p.398, l.−10–l.−9] to which Hirsch–Smale–Devaney [23, p.398, Theorem] leads fails to completely describe what it should.

(B)

Lemma. Let $f(t,y)$ be continuous on a (t,y)-set E. Let $y = y(t)$ be a solution of $y' = f(t,y)$ on $[a,\delta)$, $\delta < \infty$, for which $\exists t_n \in [a,\delta) : (\lim_{n\to\infty} t_n = \delta$ and $\lim_{n\to\infty} y(t_n) = y_0)$. If $f(t,y)$ is bounded on the intersection of E and a vicinity of the point (δ, y_0), then $\lim_{t\to\delta} y(t) = y_0$ [Hartman [40, p.13, Lemma 3.1]].

Proof. I. By hypothesis, we may take a small $\varepsilon > 0$, and a large $M_\varepsilon > 0$ such that $|f(t,y)| \le M_\varepsilon$ for $(t,y) \in E \cap \{(t,y) | 0 \le \delta - t \le \varepsilon, |y - y_0| \le \varepsilon\}$.
II. Take a large n such that $0 < \delta - t_n \le \frac{\varepsilon}{2M_\varepsilon}$ and $|y(t_n) - y_0| \le \varepsilon/2$. Then
III. $\forall t_{t_n \le t < \delta}, |y(t) - y(t_n)| < M_\varepsilon(\delta - t_n)$.

Proof. Assume that III were false. Then
$\exists t^*_{t_n \le t^* < \delta} : |y(t) - y(t_n)| \ge M_\varepsilon(\delta - t_n)$.
Let $t^1 = \min\{t \in [a,\delta) : |y(t) - y(t_n)| = M_\varepsilon(\delta - t_n)\}$. Then
1. $t_n < t^1 < \delta$.
2.

$$\begin{aligned} |y(t^1) - y(t_n)| &= M_\varepsilon(\delta - t_n) \\ &\le \varepsilon/2 \quad \text{(by II)}. \end{aligned}$$

3.

$$\begin{aligned} \forall t_{t_n \le t < \delta}, |y(t) - y_0| &\le |y(t) - y(t_n)| + |y(t_n) - y_0| \\ &< M_\varepsilon(\delta - t_n) + |y(t_n) - y_0| \quad \text{(by the definition of } t^1) \\ &\le \varepsilon \quad \text{(by II)}. \end{aligned}$$

4. $\forall t_{t_n \le t < \delta}, |y'(t)| = |f(t, y(t))| \le M_\varepsilon$.

Proof.

$$\begin{aligned}\delta - t \le \delta - t_n &\le \frac{\varepsilon}{2M_\varepsilon} \quad \text{(by II)}\\ &\le \varepsilon.\end{aligned}$$

By 3, $|y(t) - y_0| \le \varepsilon$. The result follows from I. □

5.

$$\begin{aligned}|y(t^1) - y(t_n)| &\le M_\varepsilon(t^1 - t_n) \quad \text{(by 4)}\\ &< M_\varepsilon(\delta - t_n) \quad \text{(by 1).}\end{aligned}$$

This would contradict the definition of t^1. □

□

Remark. The proof of Hartman [40, p.13, Lemma 3.1] is hard to read because all it contains is a series of formulas with little documentation.

(C) $y(t)$ tends to the boundary ∂E of E as $t \to \omega+$ [Hartman [40, p.13, l.2–l.3]].

Proof. I. Because $(b_k, y(b_k)) \notin \bar{E}_{n(k)}$, $(b_1, y(b_1)), (b_2, y(b_2)), \cdots$ is either unbounded or has a cluster point on the boundary ∂E of E [Hartman [40, p.13, l.-12–l.-11]].
II. Assume the statement "$y(t)$ tends to the boundary ∂E of E as $t \to \omega+$" were false. Then
$\exists t_n \in [a, \omega+) : \lim_{n\to\infty}(t_n, y(t_n)) = (\omega+, y_0) \in \bar{E}_m$.
1. Consequently, f is bounded on the intersection of E and a vincinity of the point $(\omega+, y_0)$. That is,
$\exists c \in [a, \omega+), M > 0 : (c \le t < \omega+) \Rightarrow |y'(t)| = |f(t, y(t)| \le M$. Thus,
$y(t)$ is uniformly continuous on $[c, \omega)$. We may define $y(\omega+) = \lim_{t\to\omega+} y(t)$. By Dugundji [24, p.302, Theorem 5.2], the extension of $y(t)$ is uniformly continuous on $[a, \omega+]$.
2. $y(t) : [a, \omega+] \to \mathbb{R}$ is differentiable at $\omega+$ and is a solution of $y'(t) = f(t, y(t))$ on $[a, \omega+]$.

Proof.

$$\begin{aligned}y(t) &= y(a) + \lim_{t\to\omega+}\int_a^t y'(s)ds \quad \text{(by definition)}\\ &= y(a) + \lim_{t\to\omega+}\int_a^t f(s, y(s))ds\\ &= y(a) + \int_a^{\omega+} f(s, y(s))ds \quad \text{(Rudin [72, p.27, Theorem 1.34]).}\end{aligned}$$

Consequently, $\forall t \in [a, \omega+], y(t) = y(a) + \int_a^t f(s, y(s))ds$ and $y'(t) = f(t, y(t))$. □

3. Since dist $(\bar{E}_m, \partial E) \ge 1/m$, by Hartman [40, p.11, Corollary 2.1], there exists a $\delta > \omega+$ such that the solution $y(t)$ on $[a, \omega+]$ can be extended to $[a, \delta]$. This would contradict the fact that $[a, \omega+)$ is the right maximal interval. □

Example 6.23. (The general method of finding a Green function's eigenfunction expansion: using symmetry)
$G(\mathbf{x},\mathbf{x}') = \frac{16\pi}{ab}\sum_{l,m=1}^{\infty}\sin(\frac{l\pi x}{a})\sin(\frac{l\pi x'}{a})\sin(\frac{m\pi y}{b})\sin(\frac{l\pi y'}{b})\frac{\sinh(K_{lm}z_<)\sinh[c-K_{lm}z_>]}{K_{lm}\sinh(K_{lm}c)}$, where $K_{lm}=\pi(l^2/a^2+m^2/b^2)^{1/2}$ [Jackson [44, p.129, (3.168)]].

Proof. I. Let $G(\mathbf{x},\mathbf{x}') = \frac{16\pi}{ab}\sum_{l,m=1}^{\infty}\sin(\frac{l\pi x}{a})\sin(\frac{l\pi x'}{a})\sin(\frac{m\pi y}{b})\sin(\frac{l\pi y'}{b})g(l,m,z,z')$ (by symmetry and a theorem similar to Coddington–Levinson [18, p.197, Theorem 4.1]).
$\nabla_{\mathbf{x}}^2 G = \frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}$
$= \frac{16\pi}{ab}\sum_{l,m=1}^{\infty}\sin(\frac{l\pi x}{a})\sin(\frac{l\pi x'}{a})\sin(\frac{m\pi y}{b})\sin(\frac{l\pi y'}{b})[\frac{\partial^2 g}{\partial z^2}-(\frac{l^2\pi^2}{a^2}+\frac{m^2\pi^2}{b^2})g]$.
$-4\pi\delta(\mathbf{x}-\mathbf{x}')$
$=-4\pi\delta(x-x')\delta(y-y')\delta(z-z')$ [Cohen-Tannoudji–Diu–Laloë [19, vol. 2, p.1477, (59)]]
$=-4\pi\delta(z-z')\sum_{l,m=1}^{\infty}\frac{4}{ab}\sin(\frac{l\pi x}{a})\sin(\frac{l\pi x'}{a})\sin(\frac{m\pi y}{b})\sin(\frac{l\pi y'}{b})$ [Cohen-Tannoudji–Diu–Laloë [19, vol. 1, p.100, (A-32)]].
Because $\nabla_{\mathbf{x}}^2 G = -4\pi\delta(\mathbf{x}-\mathbf{x}')$ [Jackson [44, p.120, (3.116)]] and $\{\sin(\frac{l\pi x}{a})\sin(\frac{m\pi y}{b})\}_{lm}$ are linearly independent,
$\frac{\partial^2 g}{\partial z^2}-K_{lm}^2 g = -\delta(z-z')$.
II. The desired result follows from Birkhoff–Rota [10, p.286, (67)]. □

Remark. (The general method of finding a Green function's eigenfunction expansion: using symmetry)

In order to reduce the problem of finding a 3-dim Green function to the problem of finding 1-dim Green function, we should summarize the proof of Jackson [44, p.121, l.2, (3.120)], the proof of Jackson [44, p.125, (3.141)], and Part I of the above proof as follows: Put the unit charge into the volume of interest. Let $\mathbf{x}'$ be its position. Let x,y,z be the Green function's three variables. Now use z' to divide the volume into two regions: I. $\{\mathbf{x}|z<z'\}$; II. $\{\mathbf{x}|z>z'\}$. In these two regions, the Poisson equation $\nabla_{\mathbf{x}}^2 G = -4\pi\delta(\mathbf{x}-\mathbf{x}')$ is reduced to the Laplace equation $\nabla_{\mathbf{x}}^2 G = 0$. Let $\{\phi_{lm}(x,y)\}_{lm}$ be the basis of the solution space. By symmetry and a theorem similar to Coddington–Levinson [18, p.197, Theorem 4.1], we have
$G(\mathbf{x},\mathbf{x}')=\sum_{lm} g_{lm}(z,z')\phi_{lm}(x,y)\phi_{lm}(x',y')$. By substituting this expression for G into $\nabla_{\mathbf{x}}^2 G = -4\pi\delta(\mathbf{x}-\mathbf{x}')$ and using Cohen-Tannoudji–Diu–Laloë [19, vol. 1, p.100, (A-32)], we will obtain the equation for 1-dim Green function.

In the last paragraph, we have used the fact that G is symmetric in (x,y) and (x',y'). In order to find the solutions of the equation for the 1-dim Green function, we should use the fact that G is symmetric in z and z'. This usage of symmetry is more subtle, more refine, and more interesting than the previous one. In view of the example given in Coddington–Levinson [18, p.222, l.9–l.14], the algebraic methods of sloving the boundary value problems such as Jackson [44, p.121, l.2, (3.120)] ($g_l(0)$ is finite; $g_l(\infty)=0$), Jackson [44, p.125, (3.141)] ($g_m(0)$ is finite; $g_m(\infty)=0$), and Birkhoff–Rota [10, p.286, Theorem 12] are essentially the same. No wonder Jackson [44, p.120, l.1–l.9] and Jackson [44, p.125, l.-11–l.-3] have similar geometric interpretations for for region I: $\{\mathbf{x}|z<z'\}$ and region II: $\{\mathbf{x}|z>z'\}$. Jackson should have quoted Birkhoff–Rota [10, p.286, Theorem 12] whenever necessary instead of repeatng its proof many times.

Example 6.24. (The Ritz method is an effective tool for studying Sturm–Liouville Problems [Fomin–Gelfand [30, pp.198–205, §41]])
I. Calculus tools for finding extrema of functions: Kaplan [47, §2.19; §2.20].
Tools in calculus of variations for finding extrema of functionals: Direct methods (the Rayleigh–Ritz method; the method of finite differences) and using Euler equations [Courant–Hilbert [22, vol.1, chap. IV, §2]].
II. Solving Sturm–Liouville Problems effectively [Fomin–Gelfand [30, pp.196–197, Remark 2]] by the Ritz

method [Fomin–Gelfand [30, p.196, Theorem]]: construct a complete sequence of functions φ_n as in Fomin–Gelfand [30, p.195, (8)]; this sequence allows us to reduce the problem of finding the minimum of the functional $J[y]$ to the problem of finding the minimum of the function $J[\alpha_1\varphi_1 + \cdots + \alpha_n\varphi_n]$ of the n variables $\alpha_1, \cdots, \alpha_n$ [Fomin–Gelfand [30, p.195, (10)]]. Thus, it suffices to calculate y_n given in Fomin–Gelfand [30, p.196, l.13–l.14] by using calculus tools for finding extrema for functions.
III. The existence of $\lambda^{(1)}$ given in Fomin–Gelfand [30, p.200, (24)] is more constructive and effective than the existence of μ_0 given in Coddington–Levinson [18, p.195, l.-9].

Explanation. (A).
1. M defined as in Fomin–Gelfand [30, p.199, l.5] can be computed by calculus.
2. For a system's solution, we may replace its function (uncountable) form $y(x)$ with its sequence (countable) form α_k as in Fomin–Gelfand [30, p.199, (18)]. Thus, $J[y]$ is transformed to $J(\alpha_1\varphi_1 + \cdots + \alpha_n\varphi_n)$, a quadratic form in $\alpha_1, \cdots, \alpha_n$. The minimum of the latter can be computered by the methods given in Kaplan [47, §2.19; §2.20].
3. Define $\lambda_n^{(1)}, y_n^{(1)} \quad (n = 1, 2, \cdots)$ as in [Fomin–Gelfand [30, p.199, l.-10–l.-7]]. Then $\lambda_{n+1}^{(1)} \le \lambda_n^{(1)}$ [Fomin–Gelfand [30, p.200, (23)]]. Define $\lambda^{(1)}$ as in Fomin–Gelfand [30, p.200, (24)]. After obtaining $\lambda_1^{(1)}, \cdots, \lambda_m^{(1)}$, we know $\lambda^{(1)}$ is between $\lambda_m^{(1)}$ and the lower bound of $\{\lambda_n^{(1)}\}$. Thus, the possible range of $\lambda^{(1)}$ is getting shorter and shorter as the process goes on. In Fomin–Gelfand [30, p.201, l.-14–p.203, l.-3], we use the method of Lagrange multipliers to obtain Fomin–Gelfand [30, p.203, (36)] and then use Fomin–Gelfand [30, p.201, Lemma 2] to prove Fomin–Gelfand [30, p.202, (32)].
(B). In contrast, $\mu_0 = \sup_{\|u\|=1} |(\mathscr{G}u, u)| \quad (u \in C \text{ on } [a, b])$ [Coddington–Levinson [18, p.195, l.2; l.-9]]. The existence of supremum is derived from reduction to absurdity [Rudin [70, p.11, l.-17–l.-16]]. We have no way to know its location on the real line. Furthermore, as we collect more elements of the index set $(u \in I)$ and find $\sup\{(\mathscr{G}u, u) | u \in I\}$, this procedure will not help narrow down the search scope of the final supremum.
Remark. Based on (A), one can easily create a effective computer program to find $\lambda^{(1)}$. However, the idea given in (B) is useless for one to find μ_0 using a computer. Mathematicians should put more effective stuff than the content given in Coddington–Levinson [18, p.194, l.-6–p.197, l.8] into mathematical textbooks.
IV. By III, $\lambda^{(1)}, \lambda^{(2)}, \cdots; y^{(1)}, y^{(2)}, \cdots$ [Fomin–Gelfand [30, §41.4)]] can be effectively calculated using the method of Lagrange multipliers, while the existence of $\mu_k \quad (k = 0, 1, 2, \cdots)$ given in Coddington–Levinson [18, p.195, l.-9–p.196, l.-2] is derived from the $(k+1)$th level of reduction to absurdity. Furthermore, that the process of finding $\mu_0, \mu_1, \cdots$ can be continued is proved by reduction to absurdity [Coddington–Levinson [18, p.197, l.1–l.7]], while that the process of constructing $\lambda^{(1)}, \lambda^{(2)}, \cdots$ can be continued because each step of the process satisfies the conditions of the method of Lagrange multipliers. □

Example 6.25. (Derivation of the equation of the vibrating membrane)

In order to effectively solve a problem, we must quickly understand the circumstance with the minimum effort, and then directly attack the heart of the matter. The local consideration given in [§6.1; `http://personal.egr.uri.edu/sadd/mce565/Ch6.pdf`] provides a simple derivation of the equation of the vibrating membrane. Newton's law is the only requirement. Considering a circular membrane with polar coordinates only complicates the circumatance [§4.3.1; `https://theses.lib.vt.edu/theses/available/etd-08022005-145837/unrestricted/Chapter4ThinPlates.pdf`].

Fomin–Gelfand [30, p.164, (48)] is derived from the viewpoint of the calculus of variations. The derivation starts with the Hamiltonian principle and ends with the Euler equation. The principle acts like an axiom and the equation acts like a theorem. The formal development makes it difficult to see the key point. The

benefit of this approach is to provide the boundary condition [Fomin–Gelfand [30, p.164, (51)]] simultaneously.
Remark. By Courant–John [21, vol. 2, p.553, (6)], $\int\int_R [\frac{\partial}{\partial x}(u_x\psi) + \frac{\partial}{\partial y}(u_y\psi)]dxdy = \int_\Gamma \frac{\partial u}{\partial n}\psi ds$ [Fomin–Gelfand [30, p.163, l.−12–l.−10]].

The global consideration given in [§Vibrating Membranes; `http://www.math.iit.edu/~fass/Notes461_Ch7Print.pdf`] increases the difficulties of the following problems:
1. Finding the tensile force F_T [p.6, l.4].
2. The balance of forces [p.7, (1)].
3. Physical explanations of the vector triple product [p.9, l.2].
4. There is no displacement u on the right-hand side [p.10, l.−2–l.−1].
The formal operations given in [p.9, l.4; (2); p.12, −1] make it difficult to see the key point.

Example 6.26. (Finding extrema with subsidiary conditions)
I. In calculus: The method of Lagrange multipliers [Reif [68, §A·10]].
II. In calculus of variations (usually consider minima): The analog of the method of Lagrange multipliers [Fomin–Gelfand [30, p.43, Theorem 1]].
III. In statistical mechanics [Reif [68, §6·8; §6·10]] (usually consider sharp maxima [Reif [68, p.202, l.8]]):
1. The method of Lagrange multipliers [Reif [68, p.229, l.−16–p.231, l.−13]].
The shortcoming: It is difficult to explain the statements given in Reif [68, p.232, l.13–l.14] in detail. Thus, we need the following more delicate methods:
2. Using the statistical trick: a rapidly increasing function multiplied by a rapidly decreasing function will produce a sharp maximum [Reif [68, p.222, l.1–p.223, l.8]].
The sharp maximum is produced by Reif [68, p.110, (3·7·14) or p.242, (7·2·15)].
3. Using the δ-function and its Fourier transform [Reif [68, p.223, l.9–p.225, l.15]].

Example 6.27. (Physics proofs vs. mathematics proofs)

A physics proof is usually intuitive; it shows how we discover a new formula. In contrast, a mathematics proof is usually abstrast; it shows how we prove it rigorously.
Example. $\lim_{g\to\infty}\int_{-\infty}^{\infty}\frac{\sin(gx)}{\pi x}f(x) = f(0)$.

A physics proof. The noddal separation π/g of $\sin(gx)$ becomes smaller and smaller as g becomes larger and larger. A small neighborhood $(x-\pi/g, x+\pi/g)$ of $x \neq 0$ contributes to the integral a period of $\sin(gx)[f(x)/x]$, which is 0, where $f(x)/x$ can be treated as a constant. In a neighborhood of $x = 0$, $f(x)$ can be considered a constant. $\int_{-\infty}^{\infty}\frac{\sin(gx)}{\pi x}f(0) = f(0)$ follows from Rudin [72, p.244, (7)]. □

A mathematics proof. The formula $\int_{-\infty}^{\infty}\frac{\sin(gx)}{\pi x}f(x) = f(0)$ can be proved using an argument similar to the one given in Rudin [72, pp.243–244, Problem 10.44]. □

Example 6.28. (Distribution theory is a new theory that we create to avoid the contradiction that the domain of a function contains a point whose function value cannot be defined)

The concept of δ generalized function originates from physics and is considered a function for many years by physicists so that some physicists think that we may treat the generalized function as a function when introducing the concept of δ generalized function, may treat it so until reaching the critical juncture, and then jump to the right track by treating it as a generalized function. Perhaps this way will allow us to avoid a contradiction. In fact, treating the generalized function as a function in the beginning has already planted the seeds of contradiction. A contradiction does not occur simply because we fail to forsee it at

that time. Actually, a contradiction must occur. The concept of generalized function is a more delicate mathematical concept than that of function. The traditional mathematical language for functions is too crude to clearly explain the concept of generalized function.
$\nabla^2(\frac{1}{|x-x'|}) = -4\pi\delta(x-x')$ [Jackson [44, p.35, l.−4, (1.31)]].

The first proof. See Pathria [61, p.501, l.3–l.11]. □

Remark. The above argument has problems. See Redžić [67, p.2, Remark †]. However, if we handle the singular point carefully, we can make the above argument work as in the second proof.

The second proof. See Cohen-Tannoudji–Diu–Laloë [19, vol. 2, p.1477, l.−5–p.1478, l.−6]. □

Remark. In order to satisfy the conditions given in Cohen-Tannoudji–Diu–Laloë [19, vol. 2, p.1477, l.−1–p.1478, l.1], we may let $g_\varepsilon(\mathbf{r}) = 1/\varepsilon(|\mathbf{r}| \leq \varepsilon)$. The second proof is not rigorous because we treat δ-function as a function rather than a generalized function [Rudin [71, p.141, l.−7–l.−3]]. We discover that behind this significant but seemingly contradictory argument, there is actually a rich, deep and refined theory. The new theory requires more delicate analysis, language and formulation so that its meaning would not be ambiguous. A generalized function may be a function or a function whose domain contains a point which does not have a well-defined function value. However, for any testing function, the generalized function must have a well-defined value. Thus, distribution theory is a new theory that we create to avoid the contradiction that the domain of a function contains a point whose function value cannot be defined. The convergence $\lim_{\varepsilon\to 0} \frac{\sin(x/\varepsilon)}{\pi x} \to \delta(x)$ [Cohen-Tannoudji–Diu–Laloë [19, vol. 2, p.1470, (10)]] should not be interpreted as the one in pointwise sense. Otherwise, we will have a contradition [Pathria [61, p.498, l.−8–l.−7]]. If we interpreted the convergence as the one in distribution sense [Rudin [71, p.146, l.9]], the previous contradiction will not occur. If we use the concept of generalized functions, the second proof actually shows that $g_\varepsilon \to -4\pi\delta$ in the distribution sense [Rudin [71, p.146, l.9]]. Thus, the second proof can be made rigorous using disbution theory, so can the third proof.

The third proof. See Jackson [44, p.35. l.1–l.−4]. □

Remark 1. Jackson's proof is correct, but is disorganized because it uses the methods of distribution theory, but fails to use the theory's terminology. A theory has its structures. Only through the use of the theory's terminology may we clarify the structure of the proof and preserve its logical rigor.
Remark 2. The above proof can be translated into the language of distribution theory as follows: Let $r_a = \sqrt{r^2+a^2}$.
$\lim_{a\to 0} < \nabla^2(\frac{1}{r_a})|\rho >= \lim_{a\to 0} \int\int\int_{|\mathbf{r}|\leq R} d^3x \nabla^2(\frac{1}{r_a})\rho(\mathbf{x})$
$= 4\pi\varepsilon_0\Phi_a(\mathbf{x}) = -4\pi\rho(0) =< -4\pi\delta_0^{(3)}|\rho >$.
Thus, $\lim_{a\to 0} \nabla^2(\frac{1}{r_a}) = -4\pi\delta_0$ [Rudin [71, p.146, l.4–l.7]].

The fourth proof. See Redžić [67, p.5, l.6–p.6, l.9]. □

Example 6.29. (How we deal with a problem that may easily cause us to commit errors)

To prove the equality given in Courant–John [21, vol. 2, p.568, l.−12–l.−11] may easily cause us to commit errors. Even worse, the situation is too confusing to allow us to locate errors. Is it because reality often goes against mathematical conventions? If so, how should we prevent an error? If we commit an error, how should we find it and then correct it?

The advantage of the method given in Courant–John [21, vol. 2, p.567, l.9–p.568, l.-8] over the direct calculation is that we need not carry out the somewhat complicated calculation of the second of u [Courant–John [21, vol. 2, p.567, l.20–l.22]]. However, proving the equality given in Courant–John [21, vol. 2, p.568, l.-12–l.-11] may easily cause one to commit errors unless one is familiar with the consequences of choosing an orientation for a curve.

Define R_n as in Courant–John [21, vol. 2, p.567, l.-6–l.-4].

Let the polar coordinates of A, B, C, and D be $(r+h,\theta), (r+h,\theta+k), (r,\theta+k)$, and (r,θ).

I. The first parameterization γ_1 of C_n:

Let $s_0 < s_4$. Define $\gamma_1 : [s_0, s_4] \to C_n$ such that $\gamma_1(s_0) = \gamma_1(s_4) = A, \gamma_1(s_1) = B, \gamma_1(s_2) = C$, and $\gamma_1(s_3) = D$.

II. The second parameterization γ_2 of C_n:

Define γ_2 such that $\gamma_2 : [\theta, \theta+k] \to \widehat{AB}, \gamma_2 : [r+h, r] \to \overline{BC}, \gamma_2 : [\theta+k, \theta] \to \widehat{CD}, \gamma_2 : [r, r+h] \to \overline{DA}$. The four segments of C_n are parameterized respectively as four different functions; it does not matter even if their domains intersect.

According to convention, the domain $[a,b]$ of a curve must satisfy $a < b$. $\gamma_2 : [r+h, r] \to \overline{BC}$ does not comply with this convention. In fact, it reverses the sense of line segment $\gamma_1[s_1, s_2]$ [Courant–John [21, vol. 1, p.334, l.-19]]. Since the principal normal is defined as the turning direction of the tangent vector, the principal normal of a point on $\gamma_2[r+h, r]$ is the opposite of the principal normal of the corresponding point on $\gamma_1[s_1, s_2]$.

III. The third parameterization γ_3 of C_n:

In such a case, the key to preventing errors is to preseve the sense of the parameterization γ_1 when parameterizing C_n. In order to fulfill this goal, all we have to do is reverse the orientation of each of the domains of the two segments of the parameterization γ_2: $\gamma_2 : [r+h, h] \to \overline{BC}, \gamma_2 : [\theta+k, \theta] \to \widehat{CD}$. The rest of the segments of the parameterization γ_2 remain the same. The parameterization so formed is called γ_3.

Based on the paramaterization γ_3, we may easily prove the equality given in Courant–John [21, vol. 2, p.568, l.-12–l.-11]. This is because a line integral is invariant under parameter changes if the orientation of the domain of a curve is preserved and also because the only segments of γ_3 whose principal normals do not point away from the origin or away from the polar axis are $\gamma_3 : [\theta, \theta+k] \to \widehat{CD}$ and $\gamma_3 : [r, r+h] \to \overline{DA}$. Since $\frac{du}{dn} = \nabla u \cdot \mathbf{n}$, the principal normals poining toward the origin and those pointing toward the polar axis contribute the two minus signs in the formula. If we unfortunately choose the parameterization γ_2, we can still make it right as long as we pay attention to the abve-mentioned remark about γ_2.

Example 6.30. (Generalized orientations)

When studying a generalized definition, we should understand its primitive version, its entire process of revolution, and the reason for the necessity of generalization. If we proceed directly toward the most general version in axiomatic approaches, its setting usually requires a more strange language and less familiar structures which may blur the essential idea, and the algorithm to check the definition usually becomes less effective. Thus, an improper approach to generalized definitions may easily lead to an empty formality and make it difficult for us to see the advantages of generalized definitions over the primitive version. Providing several non-trivial examples alone is not enough.

I. The approach given in Courant–John [21, vol. 2] aims at the origin, the insight, and the essential idea.

(1). Choosing an advantageous setting makes us easily see the entire process of revolution [Courant–John [21, vol. 2, p.575, l.-12–l.-8]].

(2). Two ordered sets of vectors in $\mathbb{R}^n$, $(\mathbf{A}_1, \cdots, \mathbf{A}_n)$ and $(\mathbf{B}_1, \cdots, \mathbf{B}_n)$, have the same orientation if and only if $[\mathbf{A}_1, \cdots, \mathbf{A}_n; \mathbf{B}_1, \cdots, \mathbf{B}_n] > 0$ [Courant–John [21, vol. 2, p.196, l.3–l.11]].

(3). Two ordered pairs of independent vectors on the tangent plane of a surface, (ξ, η) and (ξ', η'), have the

same orientation if and only if $[\xi,\eta;\xi',\eta'] > 0$ [Courant–John [21, vol. 2, p.577, (40b)]].
(4). The orientations $\Omega(\pi^*(P))$ determined by Courant–John [21, vol. 2, p.577, (40a)] from pairs of tangential vectors $\xi(P),\eta(P)$ vary continuously with P if the unit normal vector ζ given by Courant–John [21, vol. 2, p.578, (40d)] depends continuously on P [Courant–John [21, vol. 2, p.578, l.−19–l.−16]]. The positive unit normal of S^* is defined by $\Omega(\zeta,\xi,\eta) = \Omega(x,y,z)$ [Courant–John [21, vol. 2, p.579, (40i)]].
(5). S^* has the same orientation with respect to two ordered pairs of parameters (u,v) and (u',v') provided $\frac{d(u',v')}{d(u,v)} > 0$ [Courant–John [21, vol. 2, p.581, (40s)]].
(6). We use (5) [Courant–John [21, vol. 2, p.586, (41e)]] instead of the positive unit normal to generalize the concept of orientation for a surface because on a manifold in higher dimensions there is no unique normal vector or "side" of S we can associate with S [Courant–John [21, vol. 2, p.583, l.−7–p.585, l.1]].
II. In contrast, although the orientation preserving or reversing for a vector space automorphism given in Spivak [77, vol. 1, p.114, l.−5] is defined the same way as I (2), the setting for defining the concept of orientation is a non-trivial n-plane bundle [Spivak [77, vol. 1, p.116, l.−3]], a generalization of tangent bundle. The unfamiliar setting and the direct axiomatic approach to the most general version may blur the essential idea. Therefore, the approach given in Spivak [77, vol. 1, p.114, l.−8–p.118, l.−7] is definitely not suitable for beginners even though providing several non-trivial examples in the end is still not good enough.

Example 6.31. (Using formulas in a table without care may easily result in mistakes)
$\frac{1}{\pi}\int_0^\pi \frac{d\theta}{a-ib\cos\theta} = \frac{1}{\sqrt{a^2+b^2}}$, where the value of the square root is taken which makes $|a+\sqrt{a^2+b^2}| > |b|$ [Watson [89, p.384, l.6; l.12–l.13]].

Proof. Let $z = e^{i\theta}$. Then
$a - ib\cos\theta = a - \frac{ib}{2}(z+\bar{z}) = a - \frac{ib}{2}(z+\frac{1}{z}) = \frac{2az-ibz^2-ib}{2z}$.
$\int_0^\pi \frac{d\theta}{a-ib\cos\theta} = \frac{1}{2}\int_0^{2\pi}\frac{d\theta}{a-ib\cos\theta}$ [Let $f(\theta) = \frac{1}{a-ib\cos\theta}$. Then $f(\pi-\theta) = f(\pi+\theta) \Rightarrow \int_0^\pi f(\pi-\theta)d\theta = \int_0^\pi f(\pi+\theta)d\theta$]
$= \frac{1}{b}\int_{|z|=1}\frac{dz}{(z-\alpha)(z-\beta)}$ [where $\alpha,\beta = i(-\frac{a}{b} \pm \sqrt{1+\frac{a^2}{b^2}})$]
$= \frac{\pi}{\sqrt{a^2+b^2}}$ [by the residue theorem if we take the value of the square root such that $|\alpha| = |\frac{a-\sqrt{a^2+b^2}}{b}| < 1$, or equivalently $|a+\sqrt{a^2+b^2}| > |b|$ since $(a+\sqrt{a^2+b^2})(a-\sqrt{a^2+b^2}) = -b^2$]. □

Remark 1. The above argument is based on Conway [20, p.112, Example 2.9].
Remark 2. Using formulas in a table without care may easily result in mistakes. One is under the impression that once the solution form is obtained, the actual solution is determined. This is not so. If the resulting function is multivalued and the formula fails to indicate which value to choose, then the formula would be useless. One should find a delicate method to determine the correct value. If one uses such a unfinished formula in a proof, then the proof would be incorrect. Such a mistake is often difficult to detect. Here are some examples. Gradshteyn–Ryzhik [38, formula 6.611.1] fails to indicate which value of the square root to choose in order to get the correct answer. Because $(1+z)^{-1/2}$ is a multivalued function, without assigning a specific value to $(1+z)^{-1/2}$, it would be incorrect to prove $\frac{1}{a}F(\frac{1}{2},1,1,-\frac{b^2}{a^2}) = \frac{1}{\sqrt{a^2+b^2}}$ [Guo–Wang [39, p.403, (3)]] by using $(1+z)^\alpha = F(-\alpha,\beta,\beta,-z)$ [Guo–Wang [39, p.137, (10)]]. For example, one can make $(1+z)^{-1/2}$ a single-valued function by defining it as in the binormal theorem. However, one has to pay a price for doing so. For example, there are two methods to calculate $\int_0^\infty e^{-at}J_0(bt)dt$: one is using the binormal theorem to calculate $\frac{1}{a}(1+\frac{b^2}{a^2})^{-1/2}$ (one cannot calculate the square root in any other way) [Guo–Wang [39, p.403, l.11, (3)]]; the other is interpreting the square root in the answer $\frac{1}{\sqrt{a^2+b^2}}$ [Watson [89,

p.384, l.3–l.6]] in a more effective way: find two squart roots of a^2+b^2 and then choose the one satisfying $|a+\sqrt{a^2+b^2}|>|b|$.

Example 6.32. (How we detect errors in a textbook)

The formula given in Watson [89, p.388, (6)] and Guo–Wang [39, p.442, l.3] should have corrected as $\int_0^\infty e^{-t\cosh\alpha}I_\nu(t)t^{\mu-1}dt = e^{-(\mu-1/2)\pi i}\sqrt{\frac{2}{\pi}}\frac{Q^{\mu-1}_{\nu-1}(\cosh\alpha)}{(\sqrt{\sinh\alpha})^{\mu-1/2}}(*)$.

How do we detect errors in a textbook? When I find an error, the first response is usually to refuse to accept this fact and try to rationalize the opposite viewpoint. After all, there are many authors who have not found it incorrect after copying it. For example, if we replace the factor $e^{-(\mu-1/2)\pi i}$ in $(*)$ with $\frac{\cos\nu\pi}{\sin(\mu+\nu)\pi}$, then we must consider $e^{-2\pi i\mu}=1$ true. Consequently, I try to rationalize this consequence: If I properly choose the value of $\log(e^{-2\pi i})$, then $e^{-2\pi i\mu}$ can be 1. Nevertheness, I try to remember this odd experience so that I can easily find a reason when a problem occurs afterwards. However, this "rationalization" actually conceals a mistake because $e^{-2\pi i\mu}\neq 1$ if we let $\mu=1/2$. Thus, the reason why we fail to detect an error is that we have not gone far enough to forsee its consequences. Errors cannot withstand tests. Soon or later they will be detected. Even if an error may not be detected at the first checkpoint in application, it can hardly survive at the second one. When I tried to use Watson [89, p.388, (6)] and Guo–Wang [39, p.259, (4)] to prove Watson [89, p.388, (7)], I found that the coefficient of $P^{1/2-\mu}_{-\nu-1/2}$ supposed to be nonzero becomes 0 and the coefficient of $P^{1/2-\mu}_{\nu-1/2}$ supposed to be 0 becomes very complicated if we consider $\sin(\mu+\nu)\pi=\sin(\mu-\nu)\pi$ (a consequence of $e^{-2\pi i\mu}=1$) true. Thus, the world would fall into pieces as if Pandora's box were opened. I became so frustrated that I had to choose the other option: $e^{-2\pi i\mu}=1$ is not necessarily true. Then I found the counterexample: Case $\mu=1/2$. I could omit the story of proving Watson [89, p.388, (7)] and still make this paragraph logical, but this would destroy the evidence of true experience and eliminate the track of the natural thought for solving a problem.

Proof of $(*)$. (A). (Whipple's formula) $e^{-\mu\pi i}Q^\mu_\nu(\cosh\alpha)=\sqrt{\frac{\pi}{2}}\frac{\Gamma(\mu\nu+1)}{\sqrt{\sinh\alpha}}P^{-\nu-1/2}_{-\mu-1/2}(\coth\alpha)$.

Proof. I. Let $z=\cosh\alpha$; $y=P^{-\nu-1/2}_{-\mu-1/2}(w)$, where $w=\frac{z}{(z^2-1)^{1/2}}$. Then
$(1-z^2)\frac{d^2u}{dz^2}-2z\frac{du}{dz}+[\nu(\nu+1)-\frac{\mu^2}{1-z^2}]u$
$=(z^2-1)^{-5/4}\{(1-w^2)\frac{d^2y}{dw^2}-2w\frac{dy}{dw}+[(\mu^2-\frac{1}{4})-\frac{(\nu+1/2)^2}{1-w^2}]y\}=0.$
II. By I, $u(z)=AQ^\mu_\nu(z)+BP^\mu_\nu(z)$.
By Guo–Wang [39, p.249, (8); p.254, (4)], we have $B=0$ if we let ν satisfy $\Gamma(\nu+\frac{3}{2})=\infty$.
III. $\frac{Q^\mu_\nu(z)}{P^\mu_\nu(z)}\to A=e^{\mu\pi i}\Gamma(\nu+\mu+1)\sqrt{\frac{\pi}{2}}$ as $x\to+\infty$. □

(B). Let $\cosh\alpha=\coth\beta$. Then $\sinh\alpha=\operatorname{csch}\beta$.
$\int_0^\infty e^{-t\cosh\beta}I_\nu(t)t^{\mu-1}dt=\frac{\Gamma(\mu+\nu)}{\sinh^\mu\beta}P^{-\nu}_{-\mu}(\coth\beta)$ [Watson [89, p.387, (1)]; Guo–Wang [39, p.249, (9)]].
The result follows from (A). □

Example 6.33. (Contour integrals for special functions)
I. When we deal with a contour integral for a special function, all we have to do is to choose a point on the contour and assign a possible value to its argument.

The definition of contour integral for special functions is the same as that of line integral in complex

analysis [Rudin [72, p.217, (1)]]. To parameterize the integral contour for special functions, we often choose the argument of the integral's dummy variable as the parameter. Once we choose a point on the contour and assign any possible value to its argument, then the value of the integral is determined by the direction of the contour. However, branch points frequently encountered in special functions may cause a lot of confusions and complications. In order to obtain the desired solution form and facilitate the calculations for the value of the integrand near a branch point, we must choose a proper point and assign a proper value to its argument.
II. Simple notations for complicated contour integrals: Watson–Whittaker [88, p.245, l.-7–l.-5; p.256, l.-7–l.-3].
III. Confusions and complications caused by branch points:

Watson [89, p.161, l.-8–l.-7] says, "We take the phases of $t-1$ and $t+1$ to vanish at the point A." Watson–Whittaker [88, p.257, l.1] says , "At the starting-point the arguments of t and $1-t$ are both 0." We may wonder if one variable with two conditions will cause a contradiction. Guo–Wang [39, p.353, l.-7] says, "Assume $\arg(1-t^2)=0$ along the path of integration." One may wonder if this assumption is a prescription about which we should not question. Watson–Whittaker [88, p.257, l.1–l.15] shows that there are so many things concerning branch points to consider when we evaluate a contour integral. These confusions and complications are not what the authors intend to cause. The only purpose of [Watson–Whittaker [88, p.257, l.1–l.15]; Watson [89, p.161, l.-8–l.-6]; Guo–Wang [39, p.353, l.-7]] is to tell us that if we want to choose a point and its argument properly to facilitate calculations, we must consider branch points first.
IV. Convergence:

It is necessary to suppose that $\Re(\nu+\frac{1}{2})>0$ [Watson [89, p.161, l.-11–l.-10]]. This is because we must deal with branch points: the convergence of $\int_{-1}(t+1)^{\nu-1/2}dt$ or $\int^{+1}(t-1)^{\nu-1/2}dt$ requires $\Re(\nu+\frac{1}{2})>0$.
V. The advantage of representing special functions by contour integrals:

The two linearly independent solutions of the Bessel equation can be represented by the same integrand with different contours [Watson [89, p.163, (1); p.164, (2)]].

Example 6.34. (Tying up loose ends)

Both Watson [89, p.163, l.7–p.164, l.16] and Guo–Wang [39, p.355, l.1–p.356, l.-1] prove Watson [89, p.164, (3)]. However, the ways they present have shortcomings. Let us tie up loose ends. Note that if we replace $z=Re^{i\theta}$ in González [36, pp.680–681, Lemma 9.2] with $z=Re^{-i\theta}$, then the lemma will become false. If the integrand is defined as in González [36, pp.680–681, Lemma 9.2], then integral along the path $[0,\infty\exp(i\theta_1))$ equals the integral along the path $[0,\infty\exp(i\theta_2))$ if $\theta_1,\theta_2\in[0,\pi]$. For this range of available half-lines, the positive real axis is the initial half-line; the positive imaginary axis is the middle half-line; the negative real axis is the final half-line. For the above reason, we may replace the integral path $\int_0^\infty$ with $\int_0^{\infty\exp(i\theta)}$ $(0\le\theta\le\pi)$ or replace $\int_0^{i\infty}$ with $\int_0^{i\infty\exp(-i\theta)}$ $(|\theta|\le\frac{\pi}{2})$ if $|\arg z|<\frac{\pi}{2}$ [Guo–Wang [39, p.356, l.-7–l.-2]; Watson [89, p.164, l.4– l.14]]. Thus, the range of θ depends on which reference half-line we choose. For the notation $i\infty\exp(-i\omega)$, $[0,i\infty)$ represents our reference half-line and ω is used to satisfy the condition $|\arg z-\omega|<\frac{\pi}{2}$. By taking $z_0\in\{z||\arg z|<\frac{\pi}{2}\}\cap\{z||\arg z-\omega|<\frac{\pi}{2}\}(|\omega|<\frac{\pi}{2})$, we may extend the domain of z from $\{z||\arg z|<\frac{\pi}{2}\}$ to $\{z||\arg z|<\frac{\pi}{2}\}\cup\{z||\arg z-\omega|<\frac{\pi}{2}\}$ [Watson [89, p.164, l.4–l.14]]. For example, if we let $z_0\to$ the positive imaginary axis and let $\omega\to\frac{\pi}{2}$, we may extend the range of z from $\{z||\arg z|<\frac{\pi}{2}\}$ to $\{z|-\frac{\pi}{2}<\arg z<\pi\}$; if we let $z_0\to$ the negative imaginary axis and let $\omega\to-\frac{\pi}{2}$, we may extend the range of z from $\{z||\arg z|<\frac{\pi}{2}\}$ to $\{z|-\pi<\arg z<\frac{\pi}{2}\}$. In the former case, $\int_0^{i\infty\exp(i\theta)}=\int_0^{i\infty}(-\frac{\pi}{2}\le\theta\le\pi)$; in the latter case, $\int_0^{i\infty\exp(i\theta)}=\int_0^{i\infty}(-\pi\le\theta\le\frac{1}{2}\pi)$.Thus, we will have more available half-lines along which the integrals equal the integral along the original positive imaginary axis.

Example 6.35. (The finishing touch)

Providing a solution to a problem alone is not enough; the author should tell the readers from where

the solution comes. This way can bring the readers to an advantageous point for a bird's-eye view of the circumstance. By substitution,
$(z-a)^{\alpha}(z-b)^{\beta}(z-c)^{\gamma}\int_C(t-a)^{\beta+\gamma+\alpha'-1}(t-b)^{\gamma+\alpha+\beta'-1}(t-c)^{\alpha+\beta+\gamma'-1}(z-t)^{-\alpha-\beta-\gamma}dt$ [Watson–Whittaker [88, p.292, l.−14]] is a solution of Riemann's differential equation [Watson–Whittaker [88, p.283, l.13–l.16]]. Watson–Whittaker [88, p.293, l.−19–l.−14] uses the definition of the beta function and the binomial theorem to prove that the given integral form is a solution, but Watson–Whittaker [88, §14·6] fails to explain from where the form comes. In contrast, Lebedev [52, p.239, (9.1.3); (9.1.4)] indicate that the integral form is built by means of the definition of beta function and the binomial theorem. The approach given in Guo–Wang [39, §4.5] is even better because Guo–Wang [39, p.150, (3)] comes from Guo–Wang [39, §2.14, (23)]. Guo–Wang [39, §2.14] shows that the Euler transform is an important tool for solving a differential equation of Fuchsian type with three singularities.
Remark. Guo–Wang [39, p.150, l.−4–p.151, l.3] gives a detailed calculations to find M in Lebedev [52, p.240, l.4–l.5].

Example 6.36. (Musket to kill a butterfly)
The differentiation under the integral sign for $\int_0^\infty \frac{tdt}{(z^2+t^2)(e^{2\pi t}-1)}$ [Watson–Whittaker [88, p.250, l.−4]] ($\int_0^\infty \frac{\arctan(t/z)}{e^{2\pi t}-1}dt$ resp. [Watson–Whittaker [88, p.250, l.−2]]) can be justified by either the classical method [Titchmarsh [80, p.100, l.7–l.13]] or the modern method [Rudin [72, p.246, Exercise 16]]. For the latter method, we let $X=[0,\infty), d\mu=t(e^{2\pi t}-1)^{-1}dt$, and $\varphi(z,t)=(z^2+t^2)^{-1}$ for $\int_0^\infty \frac{tdt}{(z^2+t^2)(e^{2\pi t}-1)}$; let $X=[0,\infty), d\mu=t(e^{2\pi t}-1)^{-1}dt$, and $\varphi(z,t)=\frac{\arctan(t/z)}{t}$ for $\int_0^\infty \frac{\arctan(t/z)}{e^{2\pi t}-1}dt$. From the hindsight, the uniform convergence of the infinite integral in Titchmarsh [80, p.100, l.12] hints the boundedness of φ in Rudin [72, p.246, Exercise 16] for most cases.
Remark. The modern method attacks directly toward the goal by using therems flexibly. A complex measure need not distinguish a compact integral contour from a noncompact one. A single proof is good enough for dealing with both compact and noncompact cases. Furthermore, the proof is free from complex analysis except for using the definition of analytic functions. In contrast, the classical method must follow a formal, tedious, and inflexible procedure. In order to ensure the finiteness of a contour integral, the Borel measure [Rudin [72, p.49, l.10]] on $[0,\infty)$ must distinguish a compact integral contour from a noncompact one. In fact, in order to include the case of noncompact integral contour, the modification and supplement have to use almost all the theorems in complex analysis [Titchmarsh [80, §2.8–§2.84]] and, thus, lead to unnecessary complications. The proof of Rudin [72, p.246, Exercise 16] is simpler than that of the theorem given in Titchmarsh [80, p.99, l.2–l.9] because the former proof only uses Rudin [72, p.27, Theorem 1.34] and does not involve any unnecessary fuss given in Titchmarsh [80, §2.8–§2.84]. Consequently, the modern method is better than the classical method. However, the classical method is still extensively used by modern authors [Lebedev [52, p.240, l.1]; Guo–Wang [39, p.120, l.−9; p.121, l.6]; Lang [51, chap. XII]]. Perhaps, these authors are not familiar with the modern method.

Example 6.37. (Grasping the overall situation)
$M_{k,m}(z)=z^{1/2+m}e^{-z/2}\{1+\frac{1/2+m-k}{1!(2m+1)}z+\frac{(1/2+m-k)(3/2+m-k)}{2!(2m+1)(2m+2)}z^2+\cdots\}$ and
$M_{k,-m}(z)=z^{1/2-m}e^{-z/2}\{1+\frac{1/2-m-k}{1!(1-2m)}z+\frac{(1/2-m-k)(3/2-m-k)}{2!(1-2m)(2-2m)}z^2+\cdots\}$ are two linearly independent solutions near $z=0$ of
$\frac{d^2W}{dz^2}+\{-\frac{1}{4}+\frac{k}{z}+\frac{1/4-m^2}{z^2}\}W=0$ [Watson–Whittaker [88, p.337, l.−7–p.338, l.2]].

Proof. I. $F(\alpha,\gamma,z)$ and $z^{1-\gamma}F(\alpha-\gamma+1,2-\gamma,z)$ are two linearly independent solutions near $z=0$ of
$z\frac{d^2y}{dz^2}+(\gamma-z)\frac{dy}{dz}-\alpha y=0$ [Statement: Guo–Wang [39, p.297, (1), (2) & (3)]; proof: Lebedev [52, p.262,

l.−7–p.263, l.9]].
II. By means of the transformation $y = e^{z/2}z^{-\gamma/2}w(z)$,
$z\frac{d^2y}{dz^2} + (\gamma - z)\frac{dy}{dz} - \alpha y = 0$ is transformed to
$w'' + [-\frac{1}{4} + (\gamma/2 - \alpha)\frac{1}{z} + \frac{\gamma}{2}(1-\gamma)\frac{1}{z^2}]w = 0$,
i.e., $w'' + [-\frac{1}{4} + \frac{k}{z} + \frac{1/4-m^2}{z^2}]w = 0 (\gamma = 1+2m; \gamma/2 - \alpha = k)$.
$M_{k,m}(z) = e^{-z/2}z^{1/2+m}F(1/2+m-k, 1+2m, z)$ [Guo–Wang [39, p.301, (5)]] and
$M_{k,-m}(z) = e^{-z/2}z^{1/2-m}F(1/2-m-k, 1-2m, z)$ [Guo–Wang [39, p.301, (6)]] are two linearly independent solutions near $z = 0$ of the last differential equation. □

Remark 1. Watson–Whittaker [88, p.338, l.3–l.10] gives a brief summary of the above proof. Watson–Whittaker [88, §16·1] shows that $M_{k,m}(z)$ and $M_{k,-m}(z)$ are solutions of Watson–Whittaker [88, p.337, (B)] by substitution. This approach has the shortcoming of losing the beautiful structure of solution.
Remark 2. (Grasping the overall situation)

Hypergeometric functions and confluent hypergeometric functions are closely related. We must build paths between the two topics as many as possible. When we discuss confluent hypergeometric functions, of course, we have to include their characteristic properties. Furthermore, for each property, we should find its corresponding property in hypergeometric functions, treat the latter as a motivation of the former and use the latter to prove the former. Just because of the complicated circumstance, we should give a rigorous proof rather than touch it lightly. Otherwise, the discussion is incomplete.

Sneddon [74, p.32, l.1–l.18] sets a good example for discussing confluent hypergeometric functions. It says that
I. By replacing x with x/β in Sneddon [74, p.23, (8.1)] (its formal solution is given by Watson–Whittaker [88, p.207, l.7]), $F(\alpha,\beta;\gamma;x/\beta)$ is a solution of $x(1-\frac{x}{\beta})\frac{d^2y}{dx^2} + \{\gamma - (1+\frac{\alpha+1}{\beta})x\}\frac{dy}{dx} - \alpha y = 0$.
II. By Hartman [40, pp.4–5, Theorem 2.4], $\lim_{\beta\to\infty} F(\alpha,\beta;\gamma;x/\beta)$ is a solution of $x\frac{d^2y}{dx^2} + (\gamma - x)\frac{dy}{dx} - \alpha y = 0$.
Consequently, by the uniqueness of solution, we have
III. $\lim_{\beta\to\infty} F(\alpha,\beta;\gamma;x/\beta) = \sum_{r=0}^{\infty}\frac{(\alpha)_r}{(\gamma)_r}\cdot\frac{x^r}{r!}$.

By comparison, Lebedev [52, §9.9] only mentions III. However, its proof is incorrect: "a comparison of (9.1.2) and (9.1.3)" given in Lebedev [52, p.261, l.11] should have been replaced with "a comparison of (9.1.6) and (9.11.1)". Guo–Wang [39, §6.1] only mentions the transformation from hypergeometric equation to confluent hypergeometric equation [Guo–Wang [39, p.297, l.1–l.7]]. Therefore, both discussions are incomplete. Watson–Whittaker [88, chap. XVI] is poorly written because it is almost independent of Watson–Whittaker [88, chap. XIV]. A better of III is given as follows:

Proof. For $0 \le n \le t < n+1$, let $f_\beta(t) = \frac{(\alpha)_n(\beta)_n}{n!(\gamma)_n}(\frac{x}{\beta})^n$, $g(t) = \frac{(\alpha)_n(\beta)_n}{n!(\gamma)_n}(\frac{1}{2})^n$.
Let $|x| \le R$ and $|\beta| \ge 2R$. Then
$f_\beta(t) \le g(t)$ and $\lim_{\beta\to\infty} f_\beta(t) = \frac{(\alpha)_n}{n!(\gamma)_n}x^n (0 \le n \le t < n+1)$.
Treat $\sum_{n=0}^{\infty}$ as $\int_0^{\infty}$ and apply Rudin [72, p.27, Theorem 1.34] to this case. □

$F(\alpha,\gamma,z) = e^zF(\gamma-\alpha,\gamma,-z)$ [Guo–Wang [39, p.298, (6)]] can be proved similarly.

Proof. $F(\alpha,\beta,\gamma,\frac{z}{\beta} = (1-\frac{z}{\beta})^{-\beta}F(\beta,\gamma-\alpha,\gamma,\frac{z}{z-\beta})$ [Guo–Wang [39, p.143, (10)]].
For $0 \le n \le t < n+1$, let $f_\beta(t) = \frac{(\beta)_n(\gamma-\alpha)_n}{n!(\gamma)_n}(\frac{z}{z-\beta})^n$ and $g(t) = \frac{(\beta)_n(\gamma-\alpha)_n}{n!(\gamma)_n}(\frac{1}{2})^n$. Then
$f_\beta(t) \le g(t)$ if $|z| \le R$ and $|\beta| \ge 2R$. □

Example 6.38. (Linear transformations of the hypergeometric function)
I. By comparing Watson–Whittaker [88, §14·3 & §14·4] with Guo–Wang [39, §4.3], we have the following results:
(A). The former considers the general equation of Fuchsian type having three regular singularities [Guo–Wang [39, p.68, (1)]], while the latter considers the standard hypergeometric equation [Watson–Whittaker [88, p.207, Example]]. It is sufficient for our purpose to consider the standard type. In addition, it is much simpler.
(B). The former lists 24 solutions first, and then keeps 6 of them by eliminating repetitions. Through this trial-and-error approach, Watson–Whittaker [88, §14·4] finally obtains three pairs of solutions, each pair corresponding to a regular singularity [Watson–Whittaker [88, p.286, l.−17–l.−16]]. The ineffective counting shows that we should redesign our counting plan to fit our needs. That is, we should use the correspondence between solution pairs and regular singularities as the guide to redesign our counting plan. This is exactly the approach of Guo–Wang [39, §4.3].
(C). Furthermore, Guo–Wang [39, p.141, (4) & (5)] can be derived from Guo–Wang [39, p.140, (2) & (3)] by inspection [Watson–Whittaker [88, p.207, (I) & (II)]]. We may establish a similar relationship between Guo–Wang [39, p.141, (6) & (7)] and Guo–Wang [39, p.140, (2) & (3)]. It would be more difficult to recognize the above simple relationships from Watson–Whittaker [88, §14·3 & §14·4].
II. By comparing Guo–Wang [39, §4.8] with Lebedev [52, §9.5], we have the following results:
(A). (Calculations vs. inspection) By $z' = 1 - z$, the hypergeometric equation [Lebedev [52, p.248, (9.5.4)]] is transformed to the hypergeometric equation with parameters $\alpha' = \alpha, \beta' = \beta$, and $\gamma' = 1 + \alpha + \beta - \gamma$ [Lebedev [52, p.248, l.17–l.19]]. This approach cannot find the solutions of the latter differential equation without awkward calculations. In contrast, if we express the solutions of the hypergeometric equation by Riemann's P-equation [Watson–Whittaker [88, p.206, l.−7]], then the solution of the transformed hypergeometric equation can be found by inspection [Watson–Whittaker [88, p.207, (I) & (II)]]. Through Riemann's P-equation, the transformation between two singularities [] can be viewed as the transformation between two hypergeometric equations with different parameters. The solutions there are all obtained by inspection.
(B). Lebedev [52, §9.5] shows that Lebedev [52, p.249, (9.5.8) & (9.5.9); p.250, (9.5.10)] all follow from Lebedev [52, p.249, (9.5.7); p.247, (9.5.1) & (9.5.2)]. Based on the list of linear transformations given in Lebedev [52, p.246, l.−14], the discussion given in Lebedev [52, §9.5] is complete. In contrast, Guo–Wang [39, §4.8] fails to discuss Lebedev [52, p.249, (9.5.8); p.250, (9.5.10)] and fails to establish the relationship between Guo–Wang [39, p.160, (4)] and Guo–Wang [39, p.160, (8)].
(C). The formula given in Watson–Whittaker [88, p.289, l.3–l.5] and the one given in Watson–Whittaker [88, p.291, l.3–l.5] are proved the hard way because they both use the contour integrals of Barnes' type [p.286, l.−7–p.287, l.3; p.289, l.−18–l.−17] and the residue theorem. Furthermore, some cases are difficult to handle [Watson–Whittaker [88, p.290, l.14–l.16]]. In fact, we can still prove Lebedev [52, p.247, (9.5.1) & (9.5.2); p.248, (9.5.4); pp.249–250, (9.5.7)–(9.5.10)] without using any integral representation. For example, the proof of Lebedev [52, p.247, (9.5.1)] can be replaced with the proof of Guo–Wang [39, p.143, (9)]; the proof of Lebedev [52, p.244, (9.3.4)] can be replaced with the proof given in Watson–Whittaker [88, §4·11].
Remark. Reversing the order of summation and integration can be justified by Rudin [72, p.150, Theorem 7.8] or Rudin [72, p.27, Theorem 1.34]. By proving the above statement both ways, we may see the close relationship between the Fubini theorem and Lebesgue's dominated convergence theorem.

Example 6.39. (Methodical solutions)
$W_{k,m}(z) = -\frac{1}{2\pi i}\Gamma(k+\frac{1}{2}-m)e^{-z/2}z^k\int_{\infty}^{(0+)}(-t)^{-k-1/2+m}(1+\frac{t}{z})^{k-1/2+m}e^{-t}dt$ [Watson–Whittaker [88, p.339, l.−13–l.−12]] follows from Watson–Whittaker [88, p.292, l.−15–l.−10] and Guo–Wang [39, p.95, l.−8].

Remark. (Methodical solutions) The differential equation given in Watson–Whittaker [88, p.291, l.−11–l.−7] belongs to a special type. The given solution is justified simply by substitution [Watson–Whittaker [88, p.292, l.−15–l.−10]]. We do not know from where the integrand comes. The underdeveloped solution based on guess, luck, and trial-and-error such as Watson–Whittaker [88, p.339, l.−13–l.−12] cannot be considered a methodical solution. In contrast, the integral solution given in Guo–Wang [39, p.305, l.10–l.19; §6.4] is built by a systematic method which applies to the wider class of equations of Laplacian type [Guo–Wang [39, §2.13]]. In fact, the integrand and the path of integration [Guo–Wang [39, p.302, l.4–l.13]] can be specified by the Laplace transform. Consequently, the latter solution is more methodical than the former one.

Example 6.40. (Applications of analytic continuation to the Weber–Schafheitlin integral: the right timing for a statement's appearance)

Suppose we choose the weakest possible conditions required in an argument to be our theorem's hypothesis. If the argument has used the method of analytic continuation [Rudin [72, §16.9–§16.16]] no more than once, then no confusion will occur. However, what should we do if the argument has used the method of analytic continuation more than once? Let us see the following example.

Example. Let $A(z) = \sum_{m=0}^{\infty} \frac{(-)^m (a/2)^{\alpha-\beta+\gamma+2m-1}\Gamma(2\alpha+2m)\Gamma(\alpha-\beta+\gamma+2m)}{z^{2\alpha+2m} m!\Gamma(\alpha-\beta+m+1)\Gamma(\gamma+m)\Gamma(\alpha-\beta+\gamma+m)}$;
$B(z) = \int_0^\infty \frac{J_\mu(at)J_\nu(at)}{t^{\gamma-\alpha-\beta}}dt$, $D_1 = \{z|\Re(z) > 2a\}$;
$C(z) = \frac{1}{2\pi i}\int_{-\infty i}^{\infty i} \frac{(a/2)^{\alpha-\beta+\gamma+2s-1}\Gamma(2\alpha+2s)\Gamma(\alpha-\beta+\gamma+2s)}{z^{2\alpha+2s}\Gamma(\alpha-\beta+s+1)\Gamma(\gamma+s)\Gamma(\alpha-\beta+\gamma+s)}\Gamma(-s)ds$, $D_2 = \{z||\arg z| < \pi\}$.
Watson [89, p.402, l.13–l.19] shows that (B, D_1) is an analytic continuation of A; Watson [89, p.402, l.−10–l.−4] shows that (C, D_2) is ananalytic continuation of A. Since $D_1 \subset D_2$, we can say that (B, D_2) is an analytic continuation of A. In order to establish the first analytic continuation, we must impose the condition $z \in D_1$. After establishing the second analytic continuation, we find that the condition $z \in D_1$ can be weakened to the condition $z \in D_2$. However, before we establish the second analytic continuation, there is no way to know that (B, D_2) is an analytic continuation of A. Thus, the paragraph given in Watson [89, p.402, l.−13–l.−11] has the problem with timing; we should collect enough evidence before we propose a hypothesus. Therefore, whenever we use the method of analytic contiuation, we should check and record if the change of the condition is needed so that we may easily clarify the relationship between cause and effect in the proof structure.

In fact, Watson [89, §13·4; §13·41] are self-contained, but its author has written the facts in the form of previews because of the timing problem. Every time he says that a condition ensures convergence, the readers may not be able to prove the fact at that moment, but they should be able to find the proof later in the section if they are patient enough. However, some impatient readers may think that they must find the proof somewhere else. The incorrect claim given in Guo–Wang [39, p.405, l.7–l.9] is sufficient to show that there are many people under the mistaken impression. In fact, one cannot see the convergence of $\int_0^\infty \frac{J_\mu(at)J_\nu(bt)}{t^\lambda}dt$ [Watson [89, p.399, l.2–l.5]] until one reads up to Watson [89, p.401, l.15]]. Similarly, one cannot see the convergence of $\int_0^\infty \frac{J_\mu(at)J_\nu(at)}{t^\lambda}dt(\mu-\nu$ is an odd integer; $0 > \Re(\lambda) > -1)$ [Watson [89, p.403, l.−8]] until one reads up to Watson [89, p.404, (3)]].

Remark 1. $\sum_{m=0}^{\infty} \frac{(-)^m (z/2)^{\gamma+2m-1}}{m!\Gamma(\gamma+m)} |\int_0^\infty e^{-ct} J_{\alpha-\beta}(at)t^{\alpha+\beta+2m-1}dt|$ is absolutely convergent when $|z| < c$ [Watson [89, p.399, l.−9–l.−8]].

Proof. By Watson [89, p.385, (2)], $\int_0^\infty e^{-ct} J_{\alpha-\beta}(at)t^{\alpha+\beta+2m-1}dt = O(c^{-(2\alpha+2m)})$. Then use the ratio test. □

Remark 2. We impose the condition $\Re z > 0$ [Watson [89, p.399, l.18]] because $z^{\gamma+2m+1}$ [Watson [89, p.399,

l.20]] requires the consideration of the domain of $\log z$. We impose the condition $|\Im z| < c$ to ensure the convergence of $\int^{\infty} e^{-ct} J_{\gamma-1}(zt) t^{-\lambda} dt$; see Jackson [44, p.114, (3.91)].
Let $D = \{z | \Re z > 0, |z| < c\}$, $D' = \{z | \Re z > 0, |\Im z| < c, |z| < \sqrt{a^2+c^2} - c\}$ and
$f(z) = \int_0^{\infty} e^{-ct} \frac{J_{\alpha-\beta}(at) J_{\gamma-1}(zt)}{t^{\gamma-\alpha-\beta}} dt = \sum_{m=0}^{\infty} \frac{(-)^m (z/2)^{\gamma+2m-1}}{m!\Gamma(\gamma+m)} \cdot \frac{(a/2)^{\alpha-\beta}\Gamma(2\alpha+2m)}{(a^2+c^2)^{\alpha+m}\Gamma(\alpha-\beta+1)} {}_2F_1(\alpha+m, 1/2-\beta-m; \alpha-\beta+1; \frac{a^2}{a^2+c^2})$.
Watson [89, p.399, l.-14–l.-5] shows that $f(z)$ is analytic on D. Watson [89, p.400, l.1–l.23] shows that (f, D') is an analytic continuation of (f, D). In order to prove the analyticity of f on D, we impose the condition $|z| < c$; after the establishment of the analytic continuation (f, D'), we find that the condition $|z| < c$ can be weakened to the condition $|z| < \sqrt{a^2+c^2} - c$. Thus, using the method of analytic continuation is like ascending to a higher floor: our views become broader and farther.
Remark 3. By Rudin [70, p.135, Theorem 7.11], the limit of the series when $c \to 0$ is the same as the value of the series when $c = 0$ [Watson [89, p.401, l.8–l.9]]. "Provided that the integral is convergent" [Watson [89, p.401, l.-6]] means "the condition given in Watson [89, p.399, l.3] is satisfied".
Remark 4. By Jackson [44, p.114, (3.91)], $J_{\alpha-\beta}(at), J_{\gamma-1}(at) = O(e^{at})$. In order to ensure the convergence of $\int^{\infty} e^{-zt} J_{\alpha-\beta}(at) J_{\gamma-1}(at) dt$, we impose the condition $\Re z > 2a$ [Watson [89, p.402, l.-20]].
Remark 5. If $\Re z > 0$ and $|z| < 2a$, then
$\int_0^{\infty} \frac{J_{\alpha-\beta}(at) J_{\gamma-1}(at)}{t^{\gamma-\alpha-\beta}} dt = \frac{1}{2} \sum_{m=0}^{\infty} \frac{(-)^m (a/2)^{\gamma-\alpha-\beta-m-1} z^m \Gamma(\gamma-\alpha-\beta-m)\Gamma(\alpha+m/2)}{m!\Gamma(1-\beta-m/2)\Gamma(\gamma-\alpha-m/2)\Gamma(\gamma-\beta-m/2)}$
$+ \frac{1}{2} \sum_{m=0}^{\infty} \frac{(-)^m (a/2)^{-m-1} z^{\gamma-\alpha-\beta+m} \Gamma(\alpha+\beta-\gamma-m)\Gamma(\alpha/2-\beta/2+\gamma/2-m/2)}{m!\Gamma(\alpha/2-\beta/2-\gamma/2-m/2+1)\Gamma(\beta/2+\gamma/2-\alpha/2-m/2)\Gamma(\alpha/2-\beta/2+\gamma/2-m/2)}$ [Watson [89, p.403, l.3–l.5]].

Proof. We choose Watson–Whittaker [88, p.288, l.-16–p.289, l.5] or Guo–Wang [39, p.154, Fig. 9] to be our primitive model for development. By Guo–Wang [39, p.100, (8)], $\Gamma(2\alpha+2s)$ provides the factor 2^{2s}, so does $\Gamma(\alpha-\beta+\gamma+2s)$. The numerator of the integrand given in Watson [89, p.402, l.-8] provides the factor $(a/2)^{2s}$. Consequently, instead of considering $|(-z)^s \csc s\pi|$ [Guo–Wang [39, p.155, l.10]], we should consider $|(\frac{4a}{2z})^{2s} \frac{1}{\sin s\pi}|$
$= O(\exp[(N+1/2)\cos\theta \ln|\frac{4a}{2z}|^2 - (N+1/2)\delta|\sin\theta|])$ ($\ln|\frac{2a}{z}| > 0$ because $|z| < 2a$)

$$= \begin{cases} O(\exp[-2^{-1/2}(N+1/2)\ln|\frac{4a}{2z}|^2]) & \text{if } -\pi < \theta \le -3\pi/4 \text{ or } 3\pi/4 \le \theta < \pi \\ O(\exp[-2^{-1/2}(N+1/2)]) & \text{if } -3\pi/4 \le \theta \le -\pi/2 \text{ or } \pi/2 \le \theta \le 3\pi/4 \end{cases}. \qquad \square$$

Remark 6. $z^{\gamma-\alpha-\beta} = e^{(\gamma-\alpha-\beta)\ln z}$.
$|z^{\gamma-\alpha-\beta}| = e^{\Re(\gamma-\alpha-\beta)\ln|z| - \arg z \cdot \Im(\gamma-\alpha-\beta)}$.
If $\Re(\gamma-\alpha-\beta) > 0$ and $z = c \to 0$, then $|z^{\gamma-\alpha-\beta}| = e^{\Re(\gamma-\alpha-\beta)\ln c} \to 0$ [Watson [89, p.403, (1)]].
Remark 7. "It is supposed that these relations hold down to the end of §13·41." [Watson [89, p.399, l.12]] should have been corrected as follows:
"In Watson [89, p.399, l.7–p.403, l.-9], $(\mu, \nu, \lambda) \leftrightarrow (\alpha, \beta, \gamma)$ is transfo rmed according to the relations given in Watson [89, p.399, l.9–l.11]; $\alpha = (\mu+\nu-\lambda+1)/2$. In Watson [89, p.403, l.-8–p.404, l.-7], $(\mu, \nu, \lambda) \leftrightarrow (\alpha, p, \lambda)$ is transformed according to the relations given in Watson [89, p.403, l.-6–l.-5]; $\alpha = (\mu+\nu+1)/2$."
It is really confusing to use the same notation α in the same section [Watson [89, §13·41]] to represent two different quantities. The latter α should have been replaced with another notation, for example, η.
Remark 8. Without loss of generality we may assume that $p = 0, 1, 2, \cdots$ [Watson [89, p.403, l.-8; l.-6–l.-5]].
Remark 9. Since $\Re\lambda < 0$, by Bromwich [14, p.203, l.-7–l.-5; p.204, l.-17–l.-15], both ${}_2F_1(\alpha-\frac{\lambda}{2}, -p-\frac{\lambda}{2}; \alpha-p; 1)$ and ${}_2F_1(\alpha-\frac{\lambda}{2}, p+1-\frac{\lambda}{2}; \alpha+p+1; 1)$ diverge [Watson [89, p.404, l.10–l.11]]. The following

supplements may help us understand the proof of the theorem given in Bromwich [14, p.204, l.13–l.22]:
(1). In order to obtain $\frac{a_n}{a_{n+1}} < 1 + \frac{2}{n}$ [Bromwich [14, p.34, l.-8]], we must impose the condition that σ_n's are bounded.
(2). $\sum a_n$ converges $\Leftrightarrow \lim(na_n) = 0$ [Bromwich [14, p.35, l.16–l.17]].

Proof. I. $\frac{na_n}{(n+1)a_{n+1}} = 1 + \frac{1}{n}[(\mu - 1)\frac{n}{n+1} + \frac{n}{n+1}\frac{\omega_n}{n^{\lambda-1}}]$.
II. $\Rightarrow$:
By Bromwich [14, p.35, l.12], $\mu > 1$.
By I and Bromwich [14, Art. 39, Ex. 3], $\lim(na_n) = 0$.
$\Leftarrow$:
By Bromwich [14, p.35, l.12], $\mu \leq 1$.
Case $\mu < 1$: By I, $\frac{na_n}{(n+1)a_{n+1}} \leq 1$. Hence, $na_n \nearrow$.
Case $\mu = 1$: By induction, $\sum_{m=1}^{n} a_m = O(na_n)$.
If $\sum a_n$ diverges, then there exists an subsequence n_k such that $\sum_{m=1}^{n_k} a_m \to L \neq 0$.
$\exists M > 0$: $|L(n_k a_{n_k})^{-1}| \leq M$. This contradicts $\lim(na_n) = 0$. □

(3). By Rudin [70, p.62, Theorem 3.43], the hypergeometric series given in Bromwich [14, p.35, l.-16] converges for $x = -1$, if $\gamma + 1 > \alpha + \beta$.
(4). Without imposing proper conditions, the three theorems given in Bromwich [14, p.201, l.4; l.5; l.-10] cannot be valid. However, our goal is proving the theorem given in Bromwich [14, p.204, l.13–l.22]. Consequently, all we have to do is impose some conditions so that the above three theorems are valid for the cases (1), (2), and (3) given in Bromwich [14, p.204, l.-3]. For example, the theorem given in Bromwich [14, p.201, l.5] is valid for case (3) because $|\frac{a_n}{a_{n+1}}| + \frac{D_{n+1}}{D_n} > \frac{1}{2}$ ($|\frac{a_n}{a_{n+1}}| \to 1$ and $\frac{D_{n+1}}{D_n} \to 0$ as $n \to \infty$). The proof of $\underline{\lim}\, \kappa_n > 0$ [Bromwich [14, p.201, l.-10]] can be proved as follows:

Proof. $\underline{\lim}[f(n)(1 + 2\frac{f'(n)}{f(n)} + \frac{\kappa_n}{f(n)}) - \frac{f^2(n+1)}{f(n)}] > 0$ [Bromwich [14, p.201, l.4]].
$f(n)^2 + 2f(n)f'(n) - f^2(n+1)$
$= -2\int_0^1 [f(n+x)f'(n+x) - f(n)f'(n)]dx$
$= -2\int_0^1 \frac{d}{dt}[f(n+t)f'(n+t)]dt$
$= -2\int_0^1 f(n+t)(\frac{f'^2(n+t)}{f(n+t)} + f''(n+t))dt$.
For cases (1), (2), and (3), $|\frac{f(n+t)}{f(n)}| \leq 1$. $f''(n+t), \frac{f'^2(n+t)}{f(n+t)} \to 0$ [Bromwich [14, p.201, l.-5]] as $n \to \infty$. □

(5). "$\underline{\lim}(\log n)[n\{|\frac{a_n}{a_{n+1}}|^2 - 1\} - 2]] > 2$ (convergence); $\overline{\lim}(\log n)[n\{|\frac{a_n}{a_{n+1}}|^2 - 1\} - 2]] < 2$ (divergence)" [Bromwich [14, p.202, l.5, (3)]] should have been corrected as "$\underline{\lim}(\log n)[n\{|\frac{a_n}{a_{n+1}}|^2 - 1\} - 2]] > 0$ (convergence); $\overline{\lim}(\log n)[n\{|\frac{a_n}{a_{n+1}}|^2 - 1\} - 2]] < 0$ (divergence)".
(6). If $\alpha = 0$, then $|a_m| \to L > 0$ as $m \to \infty$ [Bromwich [14, p.203, l.3–l.5]].

Proof. $|\frac{a_n}{a_{n+1}}|^2 = 1 + \frac{\omega}{n^\lambda}$.
$1 - \varepsilon n^{\delta-\lambda} \leq |\frac{a_n}{a_{n+1}}|^2 \leq 1 + \varepsilon n^{\delta-\lambda}$. Consequently,
$1 - \varepsilon \sum_{k=n}^{m} k^{\delta-\lambda} \leq \sum_{k=n}^{m} |\frac{a_n}{a_{n+1}}|^2$ [Bromwich [14, p.95, l.-9]]
$\leq 1 + \varepsilon \sum_{k=n}^{m} k^{\delta-\lambda}$. □

(7). If $\mu = 1$, then $\sum a_n$ diverges [Bromwich [14, p.204, l.−7–l.−6]].

Proof. Assume that $\sum a_n$ converges to a number L.
By induction, $\sum_{m=1}^{n} a_m = O(na_n)$.
$\exists M > 0$: $|\frac{L}{na_n}| \le M$.
Since $L/n \to 0$, $a_n \not\to 0$. ☐

(8). By induction, the sum of n terms of this series is $\frac{(2-\mu)(3-\mu)\cdots(n-\mu)}{1\cdot 2\cdot 3\cdots(n-1)}$ [Bromwich [14, p.204, l.−4–l.−3]].
(9). If $0 < \alpha \le 1$, then $1 + \frac{1-\mu}{1} + \frac{(1-\mu)(2-\mu)}{1\cdot 2} + \frac{(1-\mu)(2-\mu)(3-\mu)}{1\cdot 2\cdot 3} + \cdots$ diverges [Bromwich [14, p.204, l.−2–l.−1]].

Proof. Case $0 < \alpha < 1$: $|1 + \frac{1-\mu}{n-1}|^2 \ge 1 + \frac{2(1-\alpha)}{n-1}$.
Case $\alpha = 1$: $\arg(1 - \frac{i\beta}{m}) = -\arcsin\frac{\beta}{m} \approx -\frac{\beta}{m}$.
Consequently, $\arg[\Pi_{n+1}^{\infty}(1 - \frac{i\beta}{m})] \approx -\sum_{m=n+1}^{\infty} \frac{\beta}{m} = -\beta\cdot\infty$. ☐

Remark 10. Since $\lim_{s\to-\alpha} \frac{\Gamma(2\alpha+2s)}{\Gamma(\alpha+s-p)} = \frac{1}{2}\lim_{s\to-\alpha} \frac{2(\alpha+s)\Gamma(2\alpha+2s)}{(s+\alpha)\Gamma(\alpha+s-p)}$, the residue at $s = -\alpha$ is $(-)^p/(2a)$ [Watson [89, p.404, l.−14]].
Remark 11. By Watson [89, p.403, (2), ($\lambda = 1$)] and Guo–Wang [39, p.94, (1); p.99, (3)], $\int_0^\infty \frac{J_\mu(at)J_\nu(at)}{t}dt = \frac{2}{\pi}\frac{\sin[(\nu-\mu)\pi/2]}{\nu^2-\mu^2}$ [Watson [89, p.404, l.−5]].

Example 6.41. (Integration on a Riemann surface with branch points)

If we reduce a contour integral on a Riemann surface to an integral along a line segment, the value of the latter integral may depend on which sheet the line segment is in, while the former integral is an invariant quantity. When we reduce a contour integral on a Riemann surface to an integral along a line segment, we often have to degeneate a part of the contour to a point. In order to make the argument of points along the contour continuous and simplify the calculation of these arguments, we should restore the degeneated point to its corresponding nondegeneate part. For example, in order to prove Watson [89, p.168, (3)], we must prove that
$\int_{\infty \exp i\beta}^{(0+)} e^{-u}(-u)^{\nu-1/2}(1 + \frac{iu}{2z})^{\nu-1/2}du = [e^{-i\pi(\nu-1/2)} - e^{i\pi(\nu-1/2)}]\int_0^{\infty\exp i\beta} e^{-u}u^{\nu-1/2}(1 + \frac{iu}{2z})^{\nu-1/2}du.$

Proof. Let $I = \infty\exp i\beta, A = \delta e^{i(\beta-2\pi)}, B = \delta e^{i(\beta-\pi)}, C = \delta e^{i\beta}$; IA and CI be line segments; AB and BC are counterclockwise half-circles.
Note that IA and CI are on different sheets.
$\int_{\infty\exp i\beta}^{(0+)} = \int_{IAB} + \int_{BCI}$.
We take the argument of $-u$ in the range between $\beta - 2\pi$ and β.
$\int_{BCI} = \int_0^{\infty\exp i\beta} e^{-u}(-u)^{\nu-1/2}(1 + \frac{iu}{2z})^{\nu-1/2}du$
$= (e^{-\pi i})^{\nu-1/2}\int_0^{\infty\exp i\beta} e^{-u}u^{\nu-1/2}(1 + \frac{iu}{2z})^{\nu-1/2}du.$
$\int_{IAB} = \int_{\infty\exp i\beta}^{0} e^{-u}(-u)^{\nu-1/2}(1 + \frac{iu}{2z})^{\nu-1/2}du$
$= (e^{\pi i})^{\nu-1/2}\int_{\infty\exp i\beta}^{0} e^{-u}u^{\nu-1/2}(1 + \frac{iu}{2z})^{\nu-1/2}du$
$= -(e^{\pi i})^{\nu-1/2}\int_0^{\infty\exp i\beta} e^{-u}u^{\nu-1/2}(1 + \frac{iu}{2z})^{\nu-1/2}du.$
The ending point of the integration path $[\infty\exp i\beta, 0]$ comes from the ending point of the integration path IAB, namely, B. So the argument of u at the $u = 0$ is $\beta - \pi$. Then the argument of $-u$ at the $u = 0$ is β. Thus, $\beta - (\beta - \pi) = \pi$. ☐

Similarly, in order to prove Guo–Wang [39, p.371, (11)], we must prove the equality given in Guo–Wang [39, p.371, l.7–l.8].

Proof. Let $I = 1+i\infty, A = 1+\delta e^{-3\pi i/2}, B = 1+\delta e^{-\pi i}, C = 1+\delta e^{-\pi i/2}$; IA and CI be line segments; AB and BC are counterclockwise half-circles.
$\int_{1+i\infty}^{(0+)} = \int_{IAB} + \int_{BCI}$.
Based on the restriction given in Guo–Wang [39, p.371, l.10], at the beginning point of integration path, the argument of $t-1$ is $-3\pi/2$, while the argument of $1-t$ is $-\pi/2$, so
$\int_{IAB} = \int_{1+i\infty}^{1} e^{izt}(t^2-1)^{\nu-1/2}dt = (e^{-\pi i})^{\nu-1/2}\int_{1+i\infty}^{1} e^{izt}(1-t^2)^{\nu-1/2}dt$.
$\int_{BCI} = \int_{1}^{1+i\infty} e^{izt}(t^2-1)^{\nu-1/2}dt = -\int_{1+i\infty}^{1} e^{izt}(t^2-1)^{\nu-1/2}dt$
$= -(e^{\pi i})^{\nu-1/2}\int_{1+i\infty}^{1} e^{izt}(1-t^2)^{\nu-1/2}dt$.
This is because at the beginning point of the integration path, the argument of $t-1$ is $\pi/2$, while the argument of $1-t$ is $-\pi/2$. □

Remark. The above proof shows that $\int_{1+i\infty}^{(0+)} e^{izt}(t^2-1)^{\nu-1/2}dt = (S-T)U$, where $S = (e^{-\pi i})^{\nu-1/2}, T = (e^{\pi i})^{\nu-1/2}, U = \int_{1+i\infty}^{1} e^{izt}(1-t^2)^{\nu-1/2}dt$. If we remove the restriction given in Guo–Wang [39, p.371, l.10], say, at the beginning point of the integration path in U, we let the argument of $1-t$ be $-5\pi/2$. Then U will add a factor of $(-2\pi i)^{\nu-1/2}$, S will become $e^{\pi i(\nu-1/2)}$, and T will become $e^{3\pi i(\nu-1/2}$. Thus, no matter what value we choose for U, $(S-T)U$ is an invariant quantity.

Example 6.42. (Contour integrals for Bessel functions)
Example1. $S_{\nu,\alpha,\beta,\gamma}(\rho,t;a) = \int_0^\infty J_\alpha(\rho x)J_\beta(tx)x^{\gamma+1}dx + \frac{2}{\pi}\sin\frac{(\alpha+\beta+\gamma-2\nu)\pi}{2}K_{\nu,\alpha,\beta,\gamma}(\rho,t;a)$ [Sneddon [76, p.35, l.15–l.16, (2.2.9)]].

Proof. I. Let $C_R = \{Re^{i\theta}|0\le\theta\le\frac{\pi}{2}\}$ and $F(z) = \frac{J_\nu(az)+iY_\nu(az)}{J_\nu(az)}J_\alpha(\rho z)J_\beta(tz)z^{1+\gamma}$. We want to prove $\lim_{R\to\infty}\int_{C_R}F(z)dz = 0$.

Proof. $(|\cos(az-\frac{\nu\pi}{2}-\frac{\pi}{4})|)^{-1} = 2(|e^{i(az-\frac{\nu\pi}{2}-\frac{\pi}{4})}+e^{-i(az-\frac{\nu\pi}{2}-\frac{\pi}{4})}|)^{-1} \le 4e^{\Re(iaz)}$ (note that $\Re(iz)<0$).
By Guo–Wang [39, pp.378–379], we have
$|H_\nu^{(1)}(az)| \sim \sqrt{\frac{2}{\pi aR}}e^{\Re(iaz)}$ and $|J_\alpha(\rho z)| \le \sqrt{\frac{2}{\pi aR}}e^{\Re(i\rho z)}$.
We may assume that $\arg(iz)$ lies between $\frac{\pi}{2}+\delta$ and π. □

II. $\int_{[\infty i,0]}F(z)dz = -\int_0^\infty \frac{J_\nu(aiy)+iY_\nu(aiy)}{J_\nu(aiy)}J_\alpha(\rho iy)J_\beta(tiy)(iy)^{1+\gamma}d(iy)$.
$J_\nu(aiy)+iY_\nu(aiy) = H_\nu^{(1)}(aiy)$ [Watson [89, p.73, (1)]]
$= \frac{2}{\pi}K_\nu(ay)i^{-\nu-1}$ [Jackson [44, p.116, (3.101)]].
$J_\nu(aiy) = i^\nu I_\nu(ay)$ [Jackson [44, p.116, (3.100)]].
$-\Re(i^{-2\nu+\alpha+\beta+\gamma+1}) = -\Re[e^{\pi i/2(-2\nu+\alpha+\beta+\gamma+1)}]$
$-\Re[(\cos[(-2\nu+\alpha+\beta+\gamma)/2]+i\sin[(-2\nu+\alpha+\beta+\gamma)/2])i]$
$= \sin[(-2\nu+\alpha+\beta+\gamma)/2]$.
III. Let $\Gamma = \sum_{s=1}^p \gamma_s$. Then
$\int_\Gamma F(z)dz = -\pi i\sum_{s=1}^p \operatorname{Res}F(\lambda_s)$ [González [36, p.683, Lemma 9.4]].
$Y_\nu(a\lambda_s) = \frac{-2}{\pi a\lambda_s J_\nu'(a\lambda_s)}$ [Watson [89, p.76, l.2–l.3]]
$= \frac{2}{\pi a\lambda_s J_{\nu+1}(a\lambda_s)}$ [Watson [89, p.45, (4)]].

$\lim_{p\to\infty}\int_\Gamma \Re F(z)dz = -S_{\nu,\alpha,\beta,\gamma}(\rho,t;a)$ [Guo–Wang [39, p.422, l.4–l.10]].
IV. The desired result follows from Watson [89, p.482, l.4–l.5] and Cauchy's theorem. $\square$

Example 2. $S^*_{\nu,H,\beta,\gamma,\delta} = \int_0^\infty J_\beta(ux)J_\gamma(vx)x^{\delta+1}dx + \frac{2}{\pi}\sin\frac{(\delta+\beta+\gamma-2\nu)\pi}{2}K^*_{\nu,H,\beta,\gamma,\delta}(u,v)$ [Sneddon [76, p.35, l.−4–l.−3, (2.2.10)]].

Proof. I. Let $C_R = \{Re^{i\theta}|0\le\theta\le\frac{\pi}{2}\}$ and $F(z) = \phi(z)J_\beta(uz)J_\gamma(vz)z^{\delta+1}$, where
$\phi(z) = \frac{z\{J'_\nu(z)+iY'_\nu(z)\}+H\{J_\nu(z)+iY_\nu(z)\}}{zJ'_\nu(z)+HJ_\nu(z)}$. Then $\lim_{R\to\infty}\int_{C_R}F(z)dz = 0$.
II. $\int_{[\infty i,0]}\Re F(z)dz = \frac{2}{\pi}\sin\frac{(\delta+\beta+\gamma-2\nu)\pi}{2}K^*_{\nu,H,\beta,\gamma,\delta}(u,v)$.
III. Let $\Gamma = \sum_{s=1}^p \gamma_s$. Then
$\int_\Gamma F(z)dz = -\pi i\sum_{s=1}^p \operatorname{Res}F(\mu_s)$ [González [36, p.683, Lemma 9.4]].
$\int_\Gamma \Re F(z)dz = -2\sum_{s=1}^p \frac{J_\beta(u\mu_s)J_\gamma(v\mu_s)}{(\mu_s^2-\nu^2+H^2)J_\nu^2(\mu_s)}\mu_s^{2+\delta}(0<u<1, 0<v<1)$ [Watson [89, p.480, l.21–l.24]].

Proof. $\frac{d}{dz}(zJ'_\nu(z)+HJ_\nu(z))|_{z=\mu_s} = -\frac{H^2+\mu^2-\nu^2}{\mu_s}J_\nu(\mu_s)$.
$$\begin{vmatrix} zJ'_\nu(z)+HJ_\nu(z) & zY'_\nu(z)+HY_\nu(z) \\ \frac{d}{dz}(zJ'_\nu(z)+HJ_\nu(z)) & \frac{d}{dz}(zY'_\nu(z)+HY_\nu(z)) \end{vmatrix} = z^2\begin{vmatrix} J'_\nu(z) & Y'_\nu(z) \\ J''_\nu(z) & Y''_\nu(z)\end{vmatrix}$$
$$+Hz(\begin{vmatrix} J'_\nu(z) & Y_\nu(z) \\ J''_\nu(z) & Y'_\nu(z)\end{vmatrix} + \begin{vmatrix} J_\nu(z) & Y'_\nu(z) \\ J'_\nu(z) & Y''_\nu(z)\end{vmatrix})$$
$$+(H^2+H)\begin{vmatrix} J_\nu(z) & Y_\nu(z) \\ J'_\nu(z) & Y'_\nu(z)\end{vmatrix} = \frac{2(z^2-\nu^2+H^2)}{\pi z}$$ [Watson [89, p.76, (1), (5), & (6)]]. $\square$

IV. The desired result follows from Watson [89, p.482, l.−21–l.−19] [we must assume that $\nu \ge H$] and Cauchy's theorem. $\square$

Example 6.43. (The recurrence formulas for Neumann's polynomials)
The recurrance formulas for Neumann's polynomials given by Watson [89, p.274, (1), (2) & (3)] can be derived from
I. The relation between Bessel coefficients and Neumann's polynomials: Watson [89, p.271, (1)],
II. The Laurent series expansion for the generating function of Bessel coefficients: Watson [89, p.14, (1)], and
III. The recurrence formula for Bessel coefficients: Watson [89, p.45, (1)] (see Watson [89, p.275, l.9–l.10]).
Remark 1. (Want to prove uniform convergence when convergence is given)
$\sum a_n(\frac{z}{t})^n$ converges uniformly in $(z,t) \Leftrightarrow \sum a_nJ_n(z)O_n(t)$ converges uniformly in (z,t) [Watson [89, p.274, l.3–l.5]].

Proof. Since $|\frac{z}{t}| \le (\limsup_{n\to\infty}\sqrt[n]{|a_n|})^{-1}$, $\sum\frac{a_nz^n}{t^{n+1}}O(n^{-2})$ converges uniformly in (z,t).
$\Rightarrow$: The desired result follows from Watson [89, p.273, l.−3] and Bromwich [14, p.113, l.−2].
$\Leftarrow$: We must assume that $|t|$ is bounded.
$\sum\frac{a_nz^n}{t^{n+1}}(1-\frac{z^2-t^2}{4n})$ converges uniformly in (z,t).
$\frac{|z^2-t^2|}{4}|\sum_{n=m+1}^{m+p}\frac{a_nz^n}{nt^{n+1}}| \le t^2[1+(\limsup_{n\to\infty}\sqrt[n]{|a_n|})^{-2}](4n)^{-1}|\sum_{n=m+1}^{m+p}\frac{a_nz^n}{t^{n+1}}|$ [Bromwich [14, p.113, l.−2]]. $\square$

Remark 2. (Series rearrangement) $J_0(z)[O_0(t)-tO_1(t)]+J_1(z)[2O_1(t)-\frac{1}{2}tO_2(t)+\frac{1}{2}]$
$+\sum_{n=2}^\infty J_n(z)[2O_n(t)-\frac{tO_{n+1}(t)}{n+1}-\frac{tO_{n-1}(t)}{n-1}+\frac{2n\sin^2(n\pi/2)}{n^2-1}] = 0$ [Watson [89, p.275, l.9–l.10]].

Proof. $\sum_{n=1}^{m}[J_{n-1}(z)+J_{n+1}(z)][tO_n(t)-\cos^2(n\pi/2)]/n$
$=J_0(z)tO_1(t)+J_1(z)[tO_2(t)-1]/2$
$+\sum_{m=2}^{m-1}J_n(z)[\frac{tO_{n+1}(t)}{n+1}+\frac{tO_{n-1}(t)}{n-1}-\frac{2n\sin^2(n\pi/2)}{n^2-1}]$
$+J_m(z)\{tO_{m-1}(t)-\cos^2[(m-1)\pi/2]\}/(m-1)+J_{m+1}(z)[tO_m(t)-cos^2(m\pi)/2]/m.$ □

Remark 3. (Detailed analysis) The statements given in Watson [89, p.275, l.−3–l.−1] are oversimple explanations of a complex argument. A detailed analysis should be as follows:

Proof. By the proof of Watson [89, p.14, (1)], $\sum_{n=0}^{\infty}\varepsilon_n\cos^2(n\pi/2)\cdot J_n(z)$ converges uniformly on $|z|\le B$, where B is an arbitrary positive constant.
Let $F(n,t)=2O_n(t)-\frac{tO_{n+1}(t)}{n+1}-\frac{tO_{n-1}(t)}{n-1}+\frac{2n\sin^2(n\pi/2)}{n^2-1}$.
$J_0(z)[O_0(t)-tO_1(t)]+J_1(z)[2O_1(t)-\frac{1}{2}tO_2(t)+\frac{1}{2}]$
$+\sum_{n=2}^{\infty}J_n(z)F(n,t)$ converges uniformly in $|z|\le r<R\le|t|$ [Watson [89, p.272, l.−2–l.−1]].
Assume $\exists n\ge 2: F(n,t)\not\equiv 0$.
Let $n_0=\min\{n\ge 2|F(n,t)\not\equiv 0\}$.
($\sum_{n=0}^{\infty}a_nz^{m+n}$ converges uniformly on $|z|\le 1$, where $m\in\mathbb{N}$) $\Rightarrow$ ($\sum_{n=0}^{\infty}a_nz^n$ converges uniformly on $|z|\le 1$).
Similarly, $F(n_0,t_0)+\sum_{n=n_0+1}^{\infty}[J_n(z)/J_{n_0}(z)]F(n,t_0)$ converges to 0 uniformly on a small neighborhood of $z=0$, where t_0 satisfies $F(n_0,t_0)\ne 0$.
Since ($[|z|\to 0]\Rightarrow[|J_n(z)|\to\frac{1}{n!}(\frac{|z|}{2})^n]$) [Watson [89, p.40, (8)]], we reach a contradiction. □

Example 6.44. (Determine $\arg(1-t)$ on a contour around the branch point $t=1$)

We need a method rather than correct results. Any step coming from guess may lead to the desired result this time; it may not next time. For example, if the choice $\arg(-1)=\pi$ can lead to the desired result, we want to know why we cannot choose $\arg(-1)=-\pi$. Thus, if one provides correct results without a method, one may still make mistakes sometimes. Ten correct examples are not as good as one correct method. Only when a complete method is provided may we check if results are correct. When encountering a situation where a confusion may easily occur, we should deliberately clarify the confusion rather avoid discussing it.
Example 1. After the circuit $(1+)$, $\arg(1-t)=2\pi$ [Watson–Whittaker [88, p.257, l.1–l.2]].

Proof. $t=1+\delta\exp(is)$, where $\delta>0,-\pi\le s\le\pi$.
$1-t=-\delta\exp(is)=\delta\exp[i(s+s_0)]$.
Before the circuit $(1+)$, $s=-\pi$ and $s+s_0=\arg(1-t)=0$ [Watson–Whittaker [88, p.257, l.1]]. Hence $s_0=\pi$.
After the circuit $(1+)$, $s=\pi$. So $\arg(1-t)=s+s_0=2\pi$. □

Example 2. $H_\nu^{(1)}(z)=\frac{2(z/2)^\nu}{\Gamma(\nu+1/2)\Gamma(1/2)}\int_{1+\infty i}^{1}e^{izt}(1-t^2)^{\nu-1/2}dt$ [Watson [89, p.170, l.13]].

Proof. $H_\nu^{(1)}(z)=\frac{\Gamma(1/2-\nu)(z/2)^\nu}{\pi i\Gamma(1/2)}\int_{1+\infty i}^{1}e^{izt}(t^2-1)^{\nu-1/2}dt$ [Watson [89, p.166, (4)]]].
$\int_{1+\infty i}^{(1+)}e^{izt}(t^2-1)^{\nu-1/2}dt$
$=(1-e^{2\pi i(\nu-1/2)})\int_{1+\infty i}^{1}e^{izt}(t^2-1)^{\nu-1/2}dt.$
$1-e^{2\pi i(\nu-1/2)}=e^{\pi i(\nu-1/2)}\cdot 2i\sin[(1/2-\nu)\pi]=(-1)^{\nu-1/2}\cdot 2i\frac{\pi}{\Gamma(1/2-\nu)\Gamma(1/2+\nu)}$ [Guo–Wang [39, p.99, l.4]]. □

Remark. Every formula for $H_\nu^{(1)}(z), J_\nu(z)$ and $Y_\nu(z)$ given in Watson [89, p.170, l.−18–l.−1] should have heen added the factor 2 on the right-hand side of its equality. The step given in Guo–Wang [39, p.371, l.8] gives the result, but fails to provide a method of getting the answer. According to the tha way that the solution is approached, very likely the reasoning contains guesses, and thus may be incorrect.

Example 6.45. (Binomial series)
The classical view emphasizes the choice of principal value and the consistency with previous results: The principal value of the power of a binomial given in Hobson [41, p.325, l.1] draws my attension. I find its meaning in Hobson [41, p.234, l.1; p.269, l.13]. The principal value of the power of a binomial comes from the principal value of the logarithm [Hobson [41, p.269, l.−1]]. The principal value of z^α is defined by $z^\alpha = e^{\alpha \operatorname{Log} z}$, where Log is the principal branch of the logarithmic function. Thus, the number of values of z^α is finite if α is a rational number and is countably infinite if α is an irrational number. Hobson [41, p.269, l.−3–l.−1] shows that the binomial series $f(m)$ given in Hobson [41, p.268, l.−9] converges to the principal value of $(1+z)^m$ when m is a positive rational number p/q. When m is a positive irrational number,
$f(m) = Lf(m_r)$ [Hobson [41, p.270, l.7]]
$=$ the limit of the principal value of $(1+z)^{m_r}$
$=$ the principal value of $(1+z)^m$ [by continuity or Rudin [72, p.225, Theorem 10.18]].
The modern view emphasizes whether $\sum_{k=0}^{\infty}\binom{\alpha}{k}z^k$ is convergent and whether the cases considered are inclusive: By the ratio test [Rudin [70, pp. 57–58, Theorem 3.34], we have
Theorem 1. Suppose α is not a non-negative integer. Then the radius of convergence for $\sum_{k=0}^{\infty}\binom{\alpha}{k}z^k$ is 1.
Theorem 2. Let $C = \{z \in \mathbb{C} : |z| = 1\}$. Then
(i). $\sum_{k=0}^{\infty}\binom{\alpha}{k}z^k$ converges at all points on C if $\Re\alpha > 0$.
(ii). $\sum_{k=0}^{\infty}\binom{\alpha}{k}z^k$ diverges at all points on C if $\Re\alpha \le -1$.
(iii). If $-1 < \Re\alpha \le 0$, $\sum_{k=0}^{\infty}\binom{\alpha}{k}z^k$ diverges at $z=-1$ and converges at all other points on C.

Proof. $(-1)^k\binom{\alpha}{k} = \binom{-\alpha+k-1}{k}$. By Guo–Wang [39, p.97, (3)], we have
$\binom{\alpha}{k} = \frac{(-1)^k}{\Gamma(-\alpha)k^{1+\alpha}}[1+o(1)]$ as $k \to \infty$. Consequently,
$\frac{m}{k^{1+\Re\alpha}} \le |\binom{\alpha}{k}| \le \frac{M}{k^{1+\Re\alpha}} \quad (*)$.
(i) follows from $(*)$, by comparison with the p-series $\sum_{k=1}^{\infty} k^{-p}$, where $p = 1+\Re\alpha$.
(ii). $|\binom{\alpha}{k}| \ge \frac{m}{k^{1+\Re\alpha}} \ge m > 0$.
(iii). $(1+z)\sum_{k=0}^{n}\binom{\alpha}{k}z^k = \sum_{k=0}^{n}\binom{\alpha+1}{k}z^k + \binom{\alpha}{k}z^{n+1} \quad (**)$.
By $(*)$, $\sum_{k=0}^{\infty}|\binom{\alpha}{k}z^k|$ diverges on all points on C.
If we let $z=-1$ and replace α with $\alpha-1$ in $(**)$, we have
$\sum_{k=0}^{n}\binom{\alpha}{k}(-1)^k = \binom{\alpha-1}{n}(-1)^{n+1}$. □

Example 6.46. (Listing examples cannot be considered a proof)
Listing examples cannot be considered a proof just like a tangled ball of yarn cannot be called a piece of cloth. A professional proof must give the direction of thoughts and the key idea. We should not avoid discussing the part difficult to discribe. On the contrary, we should work harder to give it a clear explanation.
Example 1. Hobson [41, p.51, l.−21–l.−6] fails to provide a proof except stating the two formulas and providing a few examples. A professional proof should be as follows:

Proof. Let $n = 2k \quad (k = 1, 2, \cdots)$.
I. How $\frac{1}{2}C_{n/2}$ becomes $D'_{(n+2)/2}$.

Proof. $C_{n/2} = \sum \cos B$, where B is the sum of k positive angles and k negative angles.
$2\cos B \sin A_{n+1} = \sin(A_{n+1}+B) + \sin(A_{n+1}-B)$.
$A_{n+1}+B$ keeps the k negative angles in B, but cannot change the k positive angles in B into negative angles. However, $A_{n+1}-B$ can change the k positive angles in B into negative angles. Consequently, there is no coefficient $1/2$ in front of $D'_{(n+2)/2}$. □

II. How $D'_{(n+2)/2}$ becomes $\frac{1}{2}C''_{(n+2)/2}$.

Proof. $D'_{(n+2)/2} = \sum \sin B$, where B is the sum of $k+1$ positive angles and k negative angles.
$2\sin B \sin A_{n+1} = \cos(B-A_{n+1}) - \cos(B+A_{n+1})$.
The $k+1$ negative angles in $B-A_{n+1}$ contains the k negative angles in B and $(-A_{n+1})$, but not the positive angles in B. Therefore, we must add the coefficient $1/2$ in front of $C''_{(n+2)/2}$. □

□

A finite series must have the first term, the last term, and the general term. An infinite series must have the first term and the general term. Thus, the two formulas given in Hobson [41, p.107, l.2 & (7)] are not correctly presented. To figure out the general term of a series from its first few terms is an example of inductive reasoning or a conjecture, but should not be considered a proof. For a binomial coefficient, we should use its compact symbol $\binom{n}{k}$ rather than its awkard factorial form $\frac{n!}{k!(n-k)!}$ unless for the purpose of computation. The formulas given in Hobson [41, p.106, l.−8–l.−1] look messy due to the abuse of notation. Suppose n is even. Hobson [41, p.107, (7)] expresses $\cos n\theta$ as a finite series in ascending powers of sine without the highest power term. Hobson [41, p.105, (3)] expresses $(-1)^{n/2}\cos n\theta$ as a finite series in descending powers of sine without the lowest power term. Hobson [41, p.107, l.12] claims that Hobson [41, p.107, (7)] is Hobson [41, p.105, (3)], written in reverse order. How is it possible to compare two things when one of them is unknown?
Example 2.
(a). A neat presentation of Hobson [41, p.107, l.2–l.3] should be as follows:
$(p+q)_s = s!\binom{q+p}{s}$
$= s!\sum_{k=0}^{s}\binom{q}{k}\binom{p}{s-k}$ [Zhu–Vandermonde's identity]
$= \sum_{k=0}^{s}\binom{s}{k}q_k p_{s-k}$.
(b). Hobson [41, p.107, (7)] should have been corrected as follows:
When n is even,
$\cos n\theta = \sum_{s=0}^{n/2}(-1)^s \frac{n^2(n^2-2^2)\cdots(n^2-[2(s-1)]^2)}{(2s)!}\sin^{2s}\theta$.
(c). A neat presentation of Hobson [41, p.106, l.9–l.−1] should be as follows:
When n is even,
$\cos n\theta = \sum_{0\le k\le n/2}(-1)^k\binom{n}{2k}(1-\sin^2\theta)^{n/2-k}\sin^{2k}\theta$
$= \sum_{s=0}^{n/2}\sin^{2s}\theta\sum_{k=0}^{s}\binom{n}{2k}\binom{n/2-k}{s-k}$.
$\sum_{k=0}^{s}\binom{n}{2k}\binom{n/2-k}{s-k} = \frac{1}{s!}\frac{n(n-2)(n-4)\cdots(n-2s+2)}{1\cdot 3\cdot 5\cdots(2s-1)}\sum_{k=0}^{s}\binom{s}{k}(\frac{2s-1}{2})_{s-k}(\frac{n-1}{2})_k$.
Only the proof of the last equality requires the use of the factorial form of binomial coefficients for computation.
(d). Suppose n is even. Hobson [41, p.107, (7)] should have been corrected as
"$\cos n\theta = \sum_{s=0}^{n/2}(-1)^s\sin^{2s}\theta[\frac{n^2(n^2-2^2)\cdots(n^2-[2(s-1)])}{(2s)!}]$."
Hobson [41, p.105, (3)] should have been corrected as
"$(-1)^{n/2}\cos n\theta = 2^{n-1}\sin^n\theta + \sum_{r=1}^{n/2}(-1)^r\sin^{n-2r}\theta[\frac{n}{r}\binom{n-r-1}{r-1}2^{n-2r-1}]$."

By replacing r with $n/2-r$ in the general term of Hobson [41, p.105, (3)], we will obtain the general term of Hobson [41, p.107, (7)].
(e). Hobson [41, §78 & §79] expresses $\cos n\theta$ and $\sin n\theta$ as descending power series of sine. Their combinatorial proofs are tedious and annoying. If we want to express them in ascending power series of sine, all have to do is list all the terms of the descending power series and then reverse the order. However, Hobson [41, §80–§83] fails to do this simple way by repeating the same kind of tedious and annoying combinatorial proofs. Mathematics is not for killing time. We have more important things to do.
(f). (Clarification of a point of confusion)
$\cos n\theta = \sum_{s=0}^{n/2}(-1)^s \sin^{2s}\theta[\frac{n^2(n^2-2^2)\cdots(n^2-[2(s-1)])}{(2s)!}]$, where n is a positive integer [Hobson [41, p.274, l.5–l.6]].

Proof. First, let us clarify the relationships among Hobson [41, §78, (1), §79, (3), §80, (7), §214, (5)]. Hobson [41, §214, (5)] and Hobson [41, §80, (7)] are the same. Hobson [41, §80, (7)] is Hobson [41, §79, (3)], written in reverse order. Hobson [41, §79, (3)] is a special case of Hobson [41, §78, (1)] when n is even. If we replace $\cos\theta$ with $(1-\sin^2\theta)^{1/2}$ in Hobson [41, §78, (1)], after expansion and rearrangement we will obtain Hobson [41, §80, (7)] when n is even. However, the symbol n in Hobson [41, §78, (1)] can represent an odd or even integer and $\{\sin^n\theta | n \in \{0\}\cup\mathbb{N}\}$ are linearly independent, so Hobson [41, §80, (7)] is true when n is odd. □

Corollary. $\frac{d^{2k+2}}{d\theta^{2k+2}}[\cos n\theta - 1 - \sum_{s=1}^{k}(-1)^s \cos^{2s}[\frac{n^2(n^2-2^2)\cdots(n^2-[2(s-1)]^2)}{(2s)!}]]|_{\theta=0} = (-1)^{k+1}[n^2(n^2-2^2)\cdots(n^2-[2k]^2)]$.

Example 6.47. (The proof of a theorem is hidden in the application whose proof requires the use of the theorem)
Example. (Stirling's theorem)
$\ln\Gamma(\lambda+s) = (s+\lambda-\frac{1}{2})\ln s - s + \frac{1}{2}\ln(2\pi) + O(s^{-1})$ [Guo–Wang [39, p.155, (6)]].

Proof. By Watson–Whittaker [88, §13·6], we obtain
$\ln\Gamma(\lambda+s) = (s+\lambda-\frac{1}{2})\ln s - s + \frac{1}{2}\ln(2\pi) + O(s^{-1+\eta}) \quad (*)$,
where η is an arbitrary small positive number. By the existence and uniqueness of the Laurent series, we have Guo–Wang [39, p.155, (6)]. □

Remark 1. I had been unable to prove Guo–Wang [39, p.155, (6)] until I attempted to prove the formula given in Watson [89, p.225, l.-12–l.-11]. All I could do was prove $(*)$ because for most applications
$\ln\Gamma(\lambda+s) = (s+\lambda-\frac{1}{2})\ln s - s + \frac{1}{2}\ln(2\pi) + O(1) \quad (**)$ is sufficient. For example, we can use $(**)$ to prove Guo–Wang [39, p.155, (7)]. However, the proof of the formula given in Watson [89, p.225, l.-12–l.-11] requires the use of Guo–Wang [39, p.155, (6)]. In this case, $(*)$ is not good enough for the high accuracy. I am able to complete the proof of Guo–Wang [39, p.155, (6)] because the second factor of the right-hand side of the formula given in Watson [89, p.225, l.-12–l.-11] is a Laurent series.
Remark 2. "Eq. (5) of Sec. 3.2" given in Guo–Wang [39, p.155, l.1] should have been corrected as "Eq. (5) of Sec. 3.21".

Example 6.48. (Finding the inverse function of a given analytic function with the Fourier series method)

The inverse function theorem provides the existence of inverse function, but fails to provide an algorithm to calculate it. J. Kepler proposed the question of finding the inverse function of a given analytic function. J. L. Lagrange solved it with two methods: the formal power series method [Guo–Wang [39, p.15, l.-9, Larrange's expansion formula] or Watson–Whittaker [88, §7·3–§7·32, Lagrange's theorem]] and the Fourier series method [see the section "A Fourier Sine Series Expansion and Resulting Bessel Function Representation for the Coefficients" in `http://www.murison.alpheratz.net/Maple/KeplerSolve/KeplerSolve.pdf`]. The later method is simpler than the former one.

Remark 1. Without reading the section "A Fourier Sine Series Expansion and Resulting Bessel Function Representation for the Coefficients" in `http://www.murison.alpheratz.net/Maple/KeplerSolve/KeplerSolve.pdf`, I would still be puzzling over how Bessel could have obtained the integral given in Watson [89, p.19, l.-6].

Remark 2. $B_n = -2(\varepsilon/n)J_n'(n\varepsilon)$ [Watson [89, p.6, l.-17]].

Proof. Let $-\varepsilon\cos E = \sum_{n=0}^{\infty} B_n \cos(nM)$.
$B_0 = -\frac{1}{\pi}\int_0^{\pi}\varepsilon\cos E dM$
$= -\frac{1}{\pi}\int_0^{\pi}\varepsilon\cos E(dE - \varepsilon\cos E dE)$ (because $M = E - \varepsilon\sin E$)
$= \varepsilon^2/2$.
Assume $n \neq 0$.
$B_n = -\frac{2}{\pi}\int_0^{\pi}\varepsilon\cos E\cos nM dM$
$= -\frac{2}{\pi}\int_0^{\pi}\frac{\varepsilon\sin nM}{n}\sin E dE$ [integration by parts]
$= -\frac{2\varepsilon}{n\pi}\int_0^{\pi}\sin n(E - \varepsilon\sin E)\sin E dE$
$= -2(\varepsilon/n)J_n'(n\varepsilon)$ [Watson [89, p.19, l.-6]]. □

Remark 3. The proof given in Watson–Whittaker [88, p.132, l.-7–p.133, l.2] requires the additional assumption that $\phi(z) \neq 0$ on and inside C, while the proof given in Guo–Wang [39, p.15, l.-7–p.16, l.11] does not.

Example 6.49. (Statements of a certain type have the same proof pattern)

Let L be a number and $(a_n) = (a_0, a_1, \cdots)$ be a number sequence. Then the following statements belong to the same type and have the same proof pattern.
$\{(a_n) | \lim_{n\to\infty} A_n = L\} \subsetneqq \{(a_n) | \lim_{n\to\infty}[A_0 + A_1 + \cdots + A_{n-1}]/n = L\}$
$\subsetneqq \{(a_n) | \lim_{n\to\infty}[nA_0 + (n-1)A_1 + \cdots + A_{n-1}]/[n(n+1)/2] = L\}$ [Bromwich [14, p.132, l.-1]].

Proof. I. $\lim_{n\to\infty} A_n = L \Rightarrow \lim_{n\to\infty}[A_0 + A_1 + \cdots + A_{n-1}]/n = L$.

Proof. $\frac{A_0 + A_1 + \cdots + A_{n-1}}{n} - L = \frac{(A_0 - L) + (A_1 - L) + \cdots + (A_{n-1} - L)}{n}$. □

II. $\lim_{n\to\infty}[A_0 + A_1 + \cdots + A_{n-1}]/n = L \Rightarrow \lim_{n\to\infty}[nA_0 + (n-1)A_1 + \cdots + A_{n-1}]/[n(n+1)/2] = L$.

Proof. $[nA_0 + (n-1)A_1 + \cdots + A_{n-1}]/[n(n+1)/2] - L$
$= \frac{A_0 + (A_0 + A_1) + \cdots + (A_0 + A_1 + \cdots + A_{N_0 - 2}) - [(N_0 - 1)N_0/2]L}{n(n+1)/2} + \frac{\frac{N_0}{n}(\frac{A_0 + \cdots + A_{N_0 - 1}}{N_0} - L) + \cdots + \frac{n}{n}(\frac{A_0 + A_1 + \cdots + A_{n-1}}{n} - L)}{(n+1)/2}$. □

III. Bromwich [14, p.133, Ex.4, (iv)] shows that $\{(a_n) | \lim_{n\to\infty}[A_0 + A_1 + \cdots + A_{n-1}]/n = L\} \backslash \{(a_n) | \lim_{n\to\infty} A_n = L\} \neq \varnothing$.
IV. $(a_n) = (0, 0, 1, -2, 3, -4, 5, -6, \cdots) \in \{(a_n) | \lim_{n\to\infty}[nA_0 + (n-1)A_1 + \cdots + A_{n-1}]/[n(n+1)/2] = \frac{1}{4}\} \backslash \{(a_n) | \lim_{n\to\infty}[A_0 + A_1 + \cdots + A_{n-1}]/n = \frac{1}{4}\}$. □

Remark. $\lim_{n\to\infty} A_n = L \Rightarrow \lim_{n\to\infty}[nA_0+(n-1)A_1+\cdots+A_{n-1}]/[n(n+1)/2] = L$.

Proof. $[nA_0+(n-1)A_1+\cdots+A_{n-1}]/[n(n+1)/2]-L$
$= n(A_0-L)+(n-1)(A_1-L)+\cdots+(A_n-L)/[n(n+1)/2]$. □

Example 6.50. (The Taylor series vs. the L'hopital rule in terms of convergence)
$\vartheta_3'(z) = \vartheta_3(z)[\sum_{n=1}^{\infty}\frac{2iq^{2n-1}e^{2iz}}{1+q^{2n-1}e^{2iz}} - \sum_{n=1}^{\infty}\frac{2iq^{2n-1}e^{-2iz}}{1+q^{2n-1}e^{-2iz}}]$ [Watson–Whittaker [88, p.471, l.6]].

Proof. $\vartheta_3(z) = G\prod_{n=1}^{\infty}(1+2q^{2n-1}\cos 2z+q^{4n-2})$ [Watson–Whittaker [88, p.469, l.−2]].
$\log\vartheta_3(z) = \log G+\sum_{n=1}^{\infty}\log(1+2q^{2n-1}\cos 2z+q^{4n-2})$.
By Watson–Whittaker [88, p.33, l.3–l.4], $\sum_{n=1}^{\infty}|\log(1+2q^{2n-1}\cos 2z+q^{4n-2})|$ converges uniformly on compact subsets of $\mathbb{C}\backslash\{$the zeros of $\vartheta_3(z)\}$.
$1+2q^{2n-1}\cos 2z+q^{4n-2} = (1+q^{2n-1}e^{2iz})(1+q^{2n-1}e^{-2iz})$.
The desired result follows from Rudin [72, p.230, Theorem 10.28]. □

Remark. Sometimes, only after studying advanced mathematics may we understand how we should properly deal with elementary mathematics. In order to study infinite products of analytic functions, we must master the concept of uniform convergence. Thus, it is important to see how the Taylor series and the L'hopital rule affect convergence. Among proofs for the case of point convergence, we should select the ones applicable to the case of uniform convergence. Watson–Whittaker [88, p.33, l.1–l.7] shows that the absolute convergence of $\sum\log(1+a_n)$ is equivalent to that of $\sum a_n$ using the Taylor series. The proof is applicable to the case of uniform convergence. The section "Convergence criteria" of `https://en.wikipedia.org/wiki/Infinite_product` proves that the convergence of $\sum\log(1+a_n)$ is equivalent to that of $\sum a_n$ using the L'hopital rule. The proof is not applicable to the case of uniform convergence because $\lim_{z\to a}\frac{f(z)}{g(z)} = \frac{f'(a)}{g'(a)}$ refers to a single point $z=a$.

Example 6.51. (Differentiation of a rational function whose denominator is a high power of a polynomial)
Let $u_3 = \frac{4z+10z^3+z^5}{8(1-z^2)^4}$ [Watson [89, p.226, l.−8]]. Find u_3'.

Solution. One may use the product rule $(fg)' = f'g+fg'$) and let $f = 4z+10z^3+z^5, g = (8(1-z^2)^4)^{-1}$; or use the quotient rule $(\frac{f}{g}) = \frac{gf'-fg'}{g^2}$ as follows:
$u_3' = \frac{1}{8}\frac{(1-z^2)^4(4+30z^2+5z^4)+8z(1-z^2)^3(4z+10z^3+z^5)}{(1-z^2)^8}$
$= \frac{1}{8}\frac{(1-z^2)(4+30z^2+5z^4)+8z(4z+10z^3+z^5)}{(1-z^2)^5}$. □

Remark. In the last step, one should not expand the expressions on the numerator of the previous step. Cancel the common factor of the numerator and the denominator first. This may avoid a lot of unnecessary computations. If one were to expand the expressions in the numerator of the first step, the resulting u_4 would be $\frac{64z+368z^3-1032z^5+273z^7+733z^9-381z^{11}-25z^{13}}{128(1-z^2)^{17/2}}$. The complicated expression would make it more difficult to identify it with the value given in Watson [89, p.226, l.−8].

Example 6.52. (The motive of creation and process of evolution for the method of steepest descents)

The interpretation of the method of steepest descents from the viewpoint of physics is the simplest and most direct [The principle of stationary phase: Watson [89, p.230, l.15–l.18]]. The interpretation from the viewpoint of mathematics is given by Guo–Wang [39, §7.11]. Watson [89, p.238, Fig. 16] [Case $x/\nu < 1$]

shows that the integral contour given in Watson [89, p.176, (3)] can be replaced with the steepest descent without changing the lower limit and the upper limit of the integral; Watson [89, p.239, Fig. 17] [Case $x/\nu > 1$] shows that each integral contour in Watson [89, p.178, (2) & (3)] can be replaced with the steepest descent without changing the lower limit and the upper limit of the integral; Watson [89, p.240, Fig. 18] [Case $x/\nu = 1$] shows that each integral contour in Watson [89, p.176, (3); p.178, (2) & (3)] can be replaced with the steepest descent without changing the lower limit and the upper limit of the integral. The arrows given in Watson [89, p.238, Fig. 16; p.239, Fig. 17; p.240, Fig. 18] can be explained by Guo–Wang [39, p.383, l.3–l.4; l.8–l.9]. The τ's given in Watson [89, p.238, l.4; p.239, l.11] come from Guo–Wang [39, p.384, l.6]. The following path shows the process of evolution for the method of steepest descent:
Abel's test [$f(x,y)$ decreases as x increases, the upper limit of each integral is ∞; Bromwich [14, p.434, l.$-$4–p.435, l.3]]
$\to$ [$f(x,n)$ decreases as x increases, $f(x,n) \to g(x)$ uniformly in any fixed finite interval, the upper limit of integral tends to ∞ as $n \to \infty$; Bromwich [14, p.443, l.6–l.11]]
$\to$ The extension to the case when f has a limit number of maxima and manima [The condition that f is positive and never increasing is removed; Bromwich [14, p.444, l.$-$6–l.$-$4]]
$\to$ [F is a function of bounded variation, the upper limit of integral tends to ∞ as $n \to \infty$; Watson [89, p.230, l.$-$12–l.$-$9]]
$\to$ Watson's lemma [Watson [89, p.236, l.11–l.19]; Guo–Wang [39, p.34, Watson's lemma]]
$\to$ the method of steepest descent [Guo–Wang [39, p.384, l.11–l.13]].
Remark. $\int_0^\infty \sin t \cdot t^{n-1} dt = \Gamma(n)\sin(\frac{1}{2}n\pi)$, where $0 < n < 1$ [Bromwich [14, p.447, l.10; p.474, l.11]; Watson [89, p.230, l.$-$10]].

Proof. I. Let $n > 0, \xi > 0, x = \xi + i\eta$, and $U = \int e^{-xt} t^{n-1} dt$. Then $U(x) = \Gamma(n)/x^n$.

Proof. $\frac{\partial U}{\partial \xi} = -\frac{n}{x}U, \frac{\partial U}{\partial \eta} = -\frac{in}{x}U$.
When $\eta = 0$, $U(x) = \Gamma(n)/\xi^n$.
$\Gamma(n)/x^n$ satisfies the above system of partial differential equations and Cauchy data.
By the Cauchy–Kowalevski theorem [John [46, p.74, l.4–l.5]], $U(x) = \Gamma(n)/x^n$ locally along the positive real axis.
By analytic continuation, $U(x) = \Gamma(n)/x^n$ in the half-plane $\xi > 0$. □

II. $\int_0^\infty e^{-it} t^{n-1} dt = \lim_{\varepsilon \to 0+} \int_0^\infty e^{-(\varepsilon+i)t} t^{n-1} dt$ [Bromwich [14, p.436, Ex. 2]]
$= \lim_{\varepsilon \to 0+} \frac{\Gamma(n)}{(\varepsilon+i)^n}$ [by I]
$= (\cos\frac{1}{2}\pi - i\sin\frac{1}{2}\pi)\Gamma(n)$. □

Example 6.53. (With vs without guess and check)

Proofs are used to check the truth of a statement and are not necessarily helpful to understand its meaning. For example, we can use the mathematical induction to prove $\sum_{k=1}^n k^2 = n(n+1)(2n+1)/6$, but do not know how we get this formula. The proof is independent of the theme of this formula in the same way as a quality control inspector checks only the packaging of product. This is a proof with guess and check; its analysis for the formula is shallow. We make the conclusion without enough confidence beforehand, and have to check afterwards; the guess and poor explanation lowers the quality of theory. Therefore, ideal and mature mathematical theories should gradually eliminate the guesswork in it. For example, the first answer in `https://math.stackexchange.com/questions/183316/how-to-get-to-the-formula-for-the-sum-of-squares-of-first-n-numbers` is a proof without guess and check. Its features: having a specific viewpoint; starting with a careful plan to get the answer; all the operations being in control beforehand. The second answer in `https://math.stackexchange.com/questions/183316/how-to-get-to-the-formula-for-the-sum-of-square` has some guesswork in the beginning. Hobson [41, p.48, (28) & (29)] are guessed from the case $n = 1,2,3$, and then proved by the mathematical induction, while Bromwich [14, §66 – §68] are without guess and check; their discussion starts with a fixed method, the rest of discussion is just the execution of calculations.

Example 6.54. (Infinite integrals)

Mathematics tends to be more effective in development [Bromwich [14, p.429, l.2–l.10]] the same way living beings tends to be more intelligent in evolution. The theory of infinite integrals emphasizes effective tests and evaluation.

(Abel's lemma) There are two proofs for Abel's lemma given in Bromwich [14, p.426, l.5–l.11]. Case when f is differentible: the proof given in Bromwich [14, p.426, l.12–l.-9] uses integration by parts. Case when f is non-differentible: the proof given in Bromwich [14, p.427, l.-9–p.428, l.-15] uses summation by parts. The two proofs are based on the same idea.

(Absolute convergence vs convergence for alternating series [Bromwich [14, p.50, l.-10–l.-8]]) Here are some corresponding effective tests: Bromwich [14, p.35, l.10–l.12] vs Bromwich [14, p.52, l.10–l.12]. Bromwich [14, p.51, Ex. 1] discusses only the case $0 < p \leq 1$ because the case $p \geq 1$ belogs to the topic of absolute convergence.

(Tests of convergence) The convergence of $\int_0^\infty (e^{-x} - e^{-tx}\frac{dx}{x}$ [Bromwich [14, p.424, l.-9]] follows from Bromwich [14, p.421, l.-6]. The convergence of $\int_0^\infty (Ae^{-ax} + Be^{-bx} + Ce^{-cx})\frac{dx}{x^2}$ [Bromwich [14, p.425, l.-12]] follows the same way. The trick of its evaluation is $\int_{a\delta}^\infty = \int_\delta^\infty - \int_\delta^{a\delta}$ [Bromwich [14, p.425, l.-6]]. The convergence of $\int_0^\infty \frac{dx}{x^2}[A\phi(ax) + B\phi(bx) + C\phi(cx)]$ follows from Dirichlet's test [Bromwich [14, p.430, l.-20–l.-18]] and its evaluation follows from the same trick $\int_{a\delta}^\infty = \int_\delta^\infty - \int_\delta^{a\delta}$. Bromwich [14, p.432, Ex. 2] discusses only the case $0 < p \leq 1$ [$p < 1$ should have been replaced with $p \leq 1$, see Bromwich [14, p.435, l.-1]] for convergence because the case $p > 1$ belongs to the topic of absolute convergence.

(Dirichlet's test) The statement of Dirichlet's test is given in Bromwich [14, p.430, l.13–l.15]. The proof is given in Buck [15, p.218, l.-12–l.-7]. Although the assumption of Buck [15, p.218, Theorem 17] is more restrictive, we may easily remove the restrictions [continuity of g'; $\int_c^\infty |g'| < \infty$] by considering $\int_c^\infty d|g|$. The proof given in Bromwich [14, p.430, l.-17–l.-13] and that given in Watson–Whittaker [88, p.72, l.4–l.7] are good for only the case when f is decreasing. If f is increasing, we must consider $-f$.

(Test of uniform convergence: the method of change of variable) Watson–Whittaker [88, p.72, l.-11–p.73, l.2] gives a typical example. The proof of uniform convergence for each of the three integrals given in Bromwich [14, p.436, l.6] is similar.

(Lebesgue's dominated convergence theorem) The proof of the theorem given in Bromwich [14, p.438, l.-16–l.-12] is similar to but simpler than that of Rudin [72, pp.246–247, Exercise 16]. The following statements can be considered corollaries of Lebesgue's dominated convergence theorem (LDCT) [Rudin [72, p.27, Theorem 1.34]]. For each statement, I indicate only the place where LDCT is used.

Bromwich [14, p.436, l.-16–l.-14] [Watson–Whittaker [88, p.73, 4·44, (I)]]: LDCT is used in Watson–Whittaker [88, p.73, l.-4].

Bromwich [14, p.438, l.-16–l.-12] [Watson–Whittaker [88, p.74, Corollary]]: LDCT is used in Bromwich [14, p.438, l.7] [Watson–Whittaker [88, p.73, l.-4]].

Bromwich [14, p.123, l.-22–-13] [Bromwich [14, p.124, l.-6–l.-1]] {Bromwich [14, p.438, l.-2–p.439, l.4]}: LDCT is used in Bromwich [14, p.124, l.8] [Apply Rudin [72, p.321, Lemma 15.3 (3)] to the tail factors and then apply LDCT to the first q factors] {Bromwich [14, p.439, l.9]}.

Bromwich [14, p.441, l.1] says that $\lim_{y\to\infty} J = 0$, by Bromwich [14, Art. 172, (1)]. This is because Bromwich [14, Art. 172, (1)] can be proved by LDCT. In my opinion, it is more direct to say that $\lim_{y\to\infty} J = 0$, by LDCT.
(The process of evolution for Abel's test for uniform convergence vs that for Weierstrass' test [Bromwich [14, p.443, l.3–l.5]])
Weierstrass' test [f is continuous, the upper limit of each integral is ∞; Bromwich [14, p.434, l.17–l.23]]
$\to$ Tannery's theorem [$f(x,n) \to g(x)$ uniformly in any fixed finite interval, the upper limit of integral tends to ∞ as $n \to \infty$; Bromwich [14, p.438, l.-2–p.439, l.4]]
$\to$ Lebesgue's dominated convergence theorem.
Abel's test [$f(x,y)$ decreases as x increases, the upper limit of each integral is ∞; Bromwich [14, p.434, l.-4–p.435, l.3]]
$\to$ [$f(x,n)$ decreases as x increases, $f(x,n) \to g(x)$ uniformly in any fixed finite interval, the upper limit of integral tends to ∞ as $n \to \infty$; Bromwich [14, p.443, l.6–l.11]]
$\to$ The extension to the case when f has alimit number of maxima and manima [The condition that f is positive and never increasing is removed; Bromwich [14, p.444, l.-6–l.-4]]
$\to$ [F is a function of bounded variation, the upper limit of integral tends to ∞ as $n \to \infty$; Watson [89, p.230, l.-12–l.-9]]
$\to$ Watson's lemma [Watson [89, p.236, l.11–l.19]; Guo–Wang [39, p.34, Watson's lemma]]
$\to$ the method of steepest descent [Guo–Wang [39, p.384, l.11–l.13]].
Remark 1. $\int_0^\infty [\sum\{A_1(1+ax)+A_2x\}e^{-ax}]\frac{dx}{x^2}$, where $\sum A_1 = \sum A_2 = 0$ and $\Re a, \Re b, \Re c, \cdots$ are positive or zero [Bromwich [14, p.441, Ex. 4]].

Proof. Let $J = \int_0^\infty e^{-xy}[\sum\{A_1(1+ax)+A_2x\}e^{-ax}]\frac{dx}{x^2}$. Then
J is uniformly convergent for $y \geq 0$.
$J'(y) = -\int_0^\infty e^{-xy}[\sum\{A_1(1+ax)+A_2x\}e^{-ax}]\frac{dx}{x}$. This integral converges uniformly so long as $y \geq l > 0$.
By Bromwich [14, p.441, Ex. 2], $J'(y) = \sum A_1 \log(a+y) - \sum(A_1a+A_2)\frac{1}{a+y}$.
$\lim_{y\to\infty} J = 0$. Note that $[\sum\{A_1(1+ax)+A_2x\}e^{-ax}]\frac{1}{x^2}$ is analytic at $x=0$.
$J(y) - J(\infty) = \int_\infty^y J'(y)dy = \lim_{M\to\infty}[\sum A_1(a+y)\log(a+y) - \sum(A_1a+A_2)\log(a+y)]|_M^y$
$= \sum A_1(a+y)\log(a+y) - \sum(A_1a+A_2)\log(a+y) - \sum A_1 a$ [use the Taylor series of $\log(1+\frac{a}{M})$]. □

Remark 2. J remains finite as y tends to ∞ [Bromwich [14, p.442, l.4]].

Proof. By the method of change of variable, $J'' = -\int_0^\infty \frac{x\sin(xy)}{1+x^2}dx$ converges uniformly in $y \geq l > 0$.
$\frac{d^2J}{dy^2} - J = -\frac{\pi}{2}$ [Bromwich [14, p.442, l.2]]. □

Remark. It would be difficult to prove the above statement if we were to consider J alone because $\lim_{x\to 0}\frac{\sin(xy)}{x} = y$.

Example 6.55. (A science book author should not use definitions to stop readers' questions)

For any science book, a reader should not accept a definition as a command about whose origin one should not question although it does not require a proof. An author should not give a definition without providing a reason. The definition given in Cohen-Tannoudji–Diu–Laloë [19, vol. 2, p.1476, (48)] fails to clearly explain from where it comes. In contrast, the formula given in Born–Wolf [12, p.894, l.-4] explains why we define the derivative of δ as in Born–Wolf [12, p.895, (13)]. This formula is established using the method of integration by parts. In fact, this method is the key to building distribution theory [Rudin [71, p.136, (1) & (3)]]. The problem $f(\infty)\delta(\infty,\mu) = 0$ [Born–Wolf [12, p.894, l.-4; p.895, l.1]] may be solved by restricting the testing functions f to the domain $\mathscr{D}$ [Rudin [71, p.136, l.9]].

It is special and interesting that the introduction to δ-function in the Fourier-transform form [Born–Wolf [12, p.896, (23)]] begins with the Fourier integral theorem [Born–Wolf [12, p.895, (19)]; Rudin [71, p.170, Theorem 7.7 (a)]]. This is because Born–Wolf [12] attempts to figure out a perfect definition from a big and useful theorem.

Example 6.56. (The right timing for correcting mistakes)

In physics, we study facts. Theories are nothing but tools to explain facts. When a theory fails to explain facts, it should be abandoned and eliminated. When we find a statement contradictory to facts, we should trace to the origin of mistake and rewrite the theory from there. For a system of identical particles, the formula given in Pathria–Beale [62, p.15, (21)] is incorrect because (i) the entropy is not an extensive property of the system [Pathria–Beale [62, p.16, l.-1–p.17, l.1]] and because (ii) Pathria–Beale [62, p.17, (4)] contradicts Pathria–Beale [62, p.18, (4a)]. Pathria–Beale [62, p.15, (21)] should have been corrected as Pathria–Beale [62, p.19, (1a)]. This is because Γ [Pathria–Beale [62, p.14, (20)]] should have been divided by $N!$ when the particles in the system are identical. However, Pathria–Beale [62, p.18, l.1–l.-3] fails to follow this simple and direct approach. Instead, it makes a fuss about it by showing the consequences if we were to assume Pathria–Beale [62, p.15, (21)] is true and by providing a remedy to satisfy the requirement. It is deplorable that Pathria–Beale [62, p.18, l.-9–l.-3] still fails to point out why the remedy can work. Of course, an incorrect statement will lead to a lot of junks, but we are not interested in why they are junks. The important thing is to correct mistakes as soon as they occur. Perhaps the Gibbs paradox is valuable for books about the development history of stastistical mechanics, but not for a textbook.

How do we choose the right timing for correcting mistakes? Shall we do it when we start to count the number of microstates [Pathria–Beale [62, p.10, l.8–l.9]]? No. If we were to consider a system of identical particles too early, we would encounter the difficulty given in Pathria–Beale [62, p.11, l.-3–l.-1]. Shall we correct mistakes at the position of Pathria–Beale [62, p.19, l.3–l.4]. No. If we were to consider a system of indentical particles and correct the mistake so late, then the validity of all the statements between Pathria–Beale [62, p.14, l.2] and Pathria–Beale [62, p.19, l.2] would become questionable. A textbook should not contain any incorrect statement because it is a reference book for quotation and application.

Example 6.57. (Maxwell made a contradiction compatible by changing $\nabla \times H = J_f$ to $\nabla \times H = J_f + J_d$)
The contradiction to be resolved: $\nabla \times H = J_f$ [Wangsness [86, p.348, (21-1)]] leads to Wangsness [86, p.348, (21-3)], which contradicts Wangsness [86, p.15, (1-49)].
His analysis: Because Wangsness [86, p.15, (1-49)] is a mathematical theorem, it must be true. Most likely, the problem arises from the incompleteness of the formula $\nabla \times H = J_f$.
His remedy for compatibility: Consequently, Maxwell changes it to $\nabla \times H = J_f + J_d$ [Wangsness [86, p.348, (21-4)]]. Then he uses the two formulas [Wangsness [86, p.152, (10-41); p.207, (12.19)]] to obtain $\nabla \cdot (\nabla \times H = -\frac{\partial}{\partial t}\nabla \cdot D + \nabla \cdot J_d$ [Wangsness [86, p.348, l.-1]]. The derivation of these two formulas is impeccable and we cannot do nothing about these universal principles. Then we have $J_d = \frac{\partial D}{\partial t}$ [Wangsness [86, p.349, (21-6)]]. Therefore, the formula $\nabla \times H = J_f$ should be corrected as $\nabla \times H = J_f + \frac{\partial D}{\partial t}$ [Wangsness [86, p.349, (21-7)]].
How the correction of the formula affects the results whose validity depends on the formula: Wangsness [86, p.349, l.13–l.-5].
Other evidence of the existence of displacement current: case $\rho > a$: if the displacement current did not exist, we would get a contradiction to the boundary condition for tangential components of H [Wangsness [86, p.352, l.4–l.13]]; $\rho < a$: Wangsness [86, p.352, (21.17)] agrees with the boundary condition for tangential

components of H because we include the displacement current as a source of H [Wangsness [86, p.352, l.14–p.353, l.7]].

Example 6.58. (Faraday made a contradiction compatible by changing $\nabla \times E = 0$ [static] to $\nabla \times E = -\frac{\partial B}{\partial t}$ [nonstatic])

Proof. Assume that $\nabla \times E = 0$ is applicable to the nonstatic case.
Wangsness [86, p.264, (17-3)] [by experiments]
$\Rightarrow$ Wangsness [86, p.266, (17-7)] [by Wangsness [86, p.266, (17-6)]]
$\Rightarrow$ Wangsness [86, p.266, (17-8)] [by Wangsness [86, p.251, (16-6)]].
Case of stationary media: Wangsness [86, p.266, l.-5–p.267, l.8];
Case of moving media: Wangsness [86, p.269, l.-6–p.272, l.10]]. Thus,
We have a contradiction. □

Remark. How the correction of the formula affects the results whose validity depends on the formula: I. Statonary media: Wangsness [86, p.267, l.9–l.-9]; II. Moving media: Wangsness [86, p.272, l.10–l.13]; localize the source of $\mathscr{E}'_m$ and interpret the origin of induced current: Wangsness [86, p.272, l.-6–l.-1]; from start to the equilibrium state described by Wangsness [86, p.273, (17-34)]: Wangsness [86, p.274, l.3–l.11]; locate the portions that contribute to the induced emf: Wangsness [86, p.276, l.10–l.14]; homopolar generator: Wangsness [86, p.276, l.-4–p.277, l.4; p.277, l.6–l.8].

Example 6.59. (How we should properly treat Ampèpere's law)
I. The situation: magnetostatics, idealized circuits [Wangsness [86, p.217, l.-23–l.-2]].
II. Our strategy: Like Coulomb's law discusses charges, Ampèpere's law discusses current elements [Wangsness [86, p.217, l.-2–p.218, l.4]]; see Wangsness [86, p.219, (13-6)]. This concept is used to provide a method of building a formula that will match experimental results [Wangsness [86, p.218, l.5–l.8]]. Applying Wangsness [86, p.218, (13-1)] to Wangsness [86, p.220, Figure 13-2] may quickly result in the desired direction $\hat{\rho}$ [Wangsness [86, p.221, (13-10)]] and a scalar double integral [Wangsness [86, p.221, (13-11)]]. If we use Wangsness [86, p.219, (13-6)] instead, we will not have these advantages. As for other aspects of comparison between Wangsness [86, p.218, (13-1)] and Wangsness [86, p.219, (13-6)], see Wangsness [86, §13-3].
III. The value of Ampèpere's law: This law is not a mathematical theorem derived from axioms. Instead, it is a method which describes a natural phenomenon mathematically and which provides an algorithm for calculating the magnetic force.

Example 6.60. (A more delicate and effective method provides more information)

Based on Wangsness [86, p.74, Figure 5-5], we can derive Wangsness [86, p.75, (5-28)]. However, this method gives no information about ρ_0 for which $\phi(\rho_0) = 0$. In contrast, based on Wangsness [86, p.57, Figure 3-8], we can derive Wangsness [86, p.75, (5-32)] and find $\rho_0 = (4L_2L_1)^{1/2}$. Both Wangsness [86, p.75, (5-28)] and Wangsness [86, p.75, (5-32)] lead to $\phi(\infty) = \infty$. However, the latter method uses Wangsness [86, p.75, (5-31)] to get a better estimate, so we can gain better information about ρ_0. See Wangsness [86, p.76, l.6–l.9].

References

[1] Ahlfors, L. V.: *Complex Analysis*, 2nd ed., New York: McGraw-Hill, 1966.

[2] Arfken, G. B. & Weber, H.J.: *Answers to Miscellaneous Problems, Mathematical Methods for physicists*, 5th ed., New York: Academic Press, 2001.

[3] Arnold, V. I.: *Mathematical Methods of Classical Mechanics*, 2nd ed., New York: Springer-Verlag, 1989.

[4] Arnold, V. I.: *Ordinary Differential Equations*, 3rd ed., translated from the Russian by Roger Cooke, New York: Springer-Verlag, 1992.

[5] Asmar, N. H.: *Partial Differential Equations with Fourier Series and Boundary Value Problems*, 2nd ed., Upper Saddle River, NJ: Prentice-Hall, 2005.

[6] Bell, R. J. T.: *An Elementary Treatise on Coordinate Geometry of Three Dimentions*, 3rd ed., London: Macmillan, 1963.

[7] Bendersky, M.: *The Calculus of Variations* `http://math.hunter.cuny.edu/mbenders/cofv.pdf`, 2008.

[8] Bernd: `http://www.macs.hw.ac.uk/%7Ebernd/F13YB1/odenotes5.pdf`

[9] Billingsley, P.: *Convergence of Probability Measures*, 2nd edition, New York: Wiley, 1999.

[10] Birkhoff, G. & Rota, G. C.: *Ordinary Differential Equations*, 3rd ed., New York: JohnWiley & Sons, 1978.

[11] Boole, G: *A Treatise on Differential Equations*, London: Macmillan, 1859.

[12] Born, M. & Wolf E.: *Principles of Optics*, 7th ed., New York: Pergamon, 2005.

[13] Borovkov, A. A.: *Probability Theory* (translated from the Russian by O. Borovkova), Amsterdam: Gordon and Breach Science Publishers, 1998.

[14] Bromwich, T. J. I'A: *An Introduction to the Theory of Infinite Series*, London: Macmillan, 1908.

[15] Buck, R. C.: *Advanced Calculus*, 3rd ed., New York: McGraw-Hill, 1978.

[16] Chevally, C.: *Theory of Lie groups I*, Princeton: Princeton University Press, 1946.

[17] Chung, K. L.: *A Course in Probability Theory*, 3rd ed., San Diego: Academic Press, 2001.

[18] Coddington E. A. & Levinson N.: *The Theory of Ordinary Differential Equations*, New York: McGraw-Hill, 1955.

[19] Cohen-Tannoudji, C., Diu, B. & Laloë, F.: *Quantum Mechanics*, 2 vols, New York: John Wiley & Sons, 1977.

[20] Conway, J. B.: *Functions of One Complex Variable*, New York: Springer-Verlag, 1973.

[21] Courant R. & John, F: *Introduction to Calculus and Analysis*, 2 vols., New York: John Wiley & sons, vol.1 (1965), vol.2 (1974).

[22] Courant R. & Hilbert D.: *Methods of Mathematical Physics*, 1st English ed., 2 vols., New York: Interscience Publishers, vol.1 (1953), vol.2 (1962).

[23] Devaney, R. L., Hirsch, M. W., Smale, S.: *Differential Equations, Dynamical Systems, and an Introduction to Chaos*, 3rd ed., New York: Academic Press, 2013.

[24] Dugundji, J.: *Topology*, Boston: Allyn and Bacon, 1966.

[25] Edwards, H. M.: *Fermat's Last Theorem (A Genetic Introduction to Algebraic Number Theory)*, New York: Spring-Verlag, 1977.

[26] Edwards, H. M.: *Galois Theory*, New York: Springer-Verlag, 1984.

[27] Ellison, W. & F.: *Prime Numbers*, New York: John Wiley & Sons, 1985.

[28] Fine, H. B. & Thompson H. D.: *Coordinate Geometry*, New York: The Macmillan Company, 1909.

[29] Fomenko, A. T.: *Symplectic Geometry*, New York: Gordon and Breach Science, 1988.

[30] Fomin, S. V. & Gelfand, I. M.: *Calculus of Variations*, translated and edited by R. A. Silverman, Englewood Cliffs, NJ: Prentice-Hall, 1963.

[31] Fowler, M.: `http://galileoandeinstein.physics.virginia.edu/lectures/michelson.html`, 2008.

[32] Gelbaum, B. R. & Olmsted, J. M. H.: *Counterexamples in Analysis*, Mineola, N.Y.: Dover, 2003.

[33] Gerlach, U.: `http://www.math.osu.edu/%7Egerlach.1/math/BVtypset/node87.html`

[34] Gibbon, J. D.: `http://www2.imperial.ac.uk/~jdg/AE2MAPDE.PDF`

[35] Goldstein, H.: *Classical Mechanics*, Reading, MA: Addison-Wesley, 1950.

[36] González, M. O.: *Classical Complex Analysis*, New York: Marcel Dekker, 1992.

[37] González, M. O.: *Complex Analysis (Selected Topics)*, New York: Marcel Dekker, 1992.

[38] Gradshteyn, I. S. & Ryzhik, I. M.: *Tables of Integrals, Series, and Products*, 4th edition, prepared by Yu. V. Geronimus and M. Yu. Tseytlin, corr. enlarged edition by A. Jeffrey, New York: Academic Press, 1980.

[39] Guo, D. R. & Wang, Z. X.: *Special Functions*, translated from the Chinese by D. R. Guo & X. J. Xio, Singapore: World Scientific, 1989.

[40] Hartman, P.: *Ordinary Differential Equations*, 2nd ed., Boston: Birkhäuser, 1982.

[41] Hobson, E. W.: *A Treatise on Plane Trigonometry*, 4th ed., Cambridge: Cambridge University Press, 1918.

[42] Hobson, E. W.: *The Theory of Spherical and Ellipsoidal Harmonics*, Cambridge: Cambridge University Press, 1931.

[43] Ince, E. L.: *Ordinary Differential Equations*, New York: Dover, 1956.

[44] Jackson, J. D.: *Classical Electrodynamics*, 3rd ed., New York: John Wiley, 1999.

[45] Jacobson, N.: *Lectures in Abstract Algebra*, 3 vols., Princeton: Van Nostrand, vol. 1 (1951), vol. 2 (1953), vol.3 (1964).

[46] John F.: *Partial Differential Equation*, 4th ed., New York: Springer-Verlag, 1982.

[47] Kaplan W.: *Advanced Calculus*, 5th ed., New York: Addison-Wesley, 2002.

[48] Koekoek, R.: Asymptotic expansions and Watson's lemma, `http://aw.twi.tudelft.nl/koekoek/documents/wi4006/watson.pdf`

[49] Kreyszig, E.: *Introduction to Differential Geometry and Riemannian Geometry*, Toronto: University of Toronto Press, 1968.

[50] Lang, S.: *Algebra*, Reading, MA, Addison-Wesley, 1965.

[51] Lang, S.: *Complex Analysis*, Reading, MA: Addison-Wesley, 1977.

[52] Lebedev, N.N.: *Special Functions and Their Applications*, translated from the Russian by R. A. Silverman, Englewood Cliffs, N.J.: Prectice-Hall, 1965.

[53] Lindgren, B. W.: *Statistical Theory*, 3rd ed., New York: Macmillan Publishing Co., 1976.

[54] Loéve, M: *Probability Theory*, vol. I, 4th ed., New York: Springer-Verlag , 1977.

[55] Marion, J. B. & Thornton, S. T.: *Classical Dynamics of Particles and Systems*, 5th ed., Belmont, C.A.: Brooks/Cole, 2004.

[56] Massey, W. S.: *Algebraic Topology*: An Introduction, New York: Springer-Verlag, 1989.

[57] Munkres J. R.: *Topology: A First Course*, Englewood Cliffs, NJ: Prentice-Hall, 1975.

[58] Munkres J. R.: *Topology*, 2nd ed., Englewood Cliffs, NJ: Prentice-Hall, 2000.

[59] O'Neill, B.: *Elementary Differential Geometry*, New York: Academic Press, 1966.

[60] Parzen, E.: *Modern Probability Theory and Its Applications*, New York: Wiley, (1960, 1992).

[61] Pathria, R. K.: *Statistical Mechanics*, 1st ed., New York: Pergamon, 1980.

[62] Pathria, R. K. & Beale, P. D.: *Statistical Mechanics*, 3rd ed., New York: Elsevier, 2011.

[63] Peebles, P. J. E.: *Principles of Physical Cosmology*, Princeton, N. J.: Princeton University Press, 1993.

[64] Perron, O.: Die Lehre von den Kettenbrüchen, Berlin, B.G. Teubner, 1913.

[65] Pontryagin, L. S.: *Ordinary Differential Equations*, translated from the Russian by Leonas Kacinskas and Walter B. Counts, Reading: Addison-Wesley Publishing Company, 1962.

[66] Pontryagin, L. S.: *Topological Groups*, 2nd ed., translated from the Russian by Arien Brown, New York: Gordon & Breach, 1966.

[67] Redžić, D.V.: *On the Laplacian of* $1/r$, `http://arxiv.org/pdf/1303.2567.pdf`.

[68] Reif, F.: *Statistical Thermal Physics*, New York: McGraw-Hill, 1965.

[69] Royden, H. L.: *Real Analysis*, 2nd ed., New York: The Macmillan Company, 1968.

[70] Rudin, W.: *Principles of Mathematical Analysis*, 2nd ed., New York: McGraw- Hill, 1964.

[71] Rudin, W.: *Functional Analysis*, New York: McGraw-Hill, 1973.

[72] Rudin, W.: *Real and Complex Analysis*, 2nd ed., New York: McGraw-Hill, 1974.

[73] Saks, S. & Zygmund, A.: *Analytic Functions*, translated by E. J. Scott, Monografje Matematyczne, vol. 28, 3rd ed., Warsaw, 1971.

[74] Sneddon, I. N.: *Special Functions of Mathematical Physics and Chemistry*, New York: Interscience Publishers, 1956.

[75] Sneddon, I. N.: *Elements of Partial Differential Equations*, New York: McGraw-Hall, 1957.

[76] Sneddon, I. N.: *Mixed Boundary Value Problems in Potential Theory*, New York: Wiley-Interscience, 1966.

[77] Spivak, M.: *A Comprehensive Introduction to Differential Geometry*, Vol. 1, Berkeley: Publish or Perish, 1979.

[78] Stewart, I. & Tall, D.: *Algebraic Number Theory*, 2nd ed., London: Chapman and Hall, 1987.

[79] Symon, K. R.: *Mechanics*, 3rd ed., Reading, MA: Addison-Wesley, 1971.

[80] Titchmarsh, E. C.: *The Theory of Functions*, 2nd ed., Oxford: Oxford University Press, 1939.

[81] Tkachev, V.: `http://www.mai.liu.se/~vlatk48/teaching/teaching_vt2009/lectures_uu/PDE09-03.pdf`

[82] van der Waerden, B. L.: *Modern Algebra*, 2 vols, Translated from the German by F. Blum, New York: Ungar, 1949.

[83] Wall, H. S.: *Analytic Theory of Continued Fractions*, New York: D. Van Nostrand Company, 1948.

[84] Wang, L. C.: `http://www.lcwangpress.com/physics/method-de.htm`

[85] Wang, L. C.: `http://www.lcwangpress.com/papers/absurdity.pdf`

[86] Wangsness, R. K.: *Electromagnetic Fields*, New York: John Wiley & Sons, 1986.

[87] Warner, F. W.: *Foundations of Differentiable Manifolds and Lie Groups*, Glenview, IL: Scott, Foresman, 1971.

[88] Watson, G. N. & Whittaker E. T.: *A Course of Modern Analysis*, 4th ed., Cambridge: Cambridge University Press, 1963.

[89] Watson, G. N.: *Theory of Bessel Functions*, 2nd ed., Cambridge: Cambridge University Press, 1966.

[90] Weatherburn, C. E.: *Differential Geometry of Three Dimensions*, 2 vols, Cambridge: The University Press, vol. 1 (1927); vol. 2 (1930).

[91] Zemansky, M. W. & Dittman, R. H.: *Heat and Thermodynamics*, New York: McGraw-Hill, 1981.

[92] Zygmund, A.: *Trigonometric Series*, 2 vols., 2nd ed., Cambridge: Cambridge University Press, 1959.

Mr. Li-Chung Wang is the author of the following website about the philosophy of mechanics:
http://www.lcwangpress.com/physics/main.html.
Address: 7th Floor, #21 Lane 267, Xi-zhou Street, Chungli, Taiwan, ROC.
E-mail:lcwangpress@yahoo.com