Existence and Uniqueness from the Viewpoint of Effectiveness

Li-Chung Wang

May 30, 2011

Abstract

In Section 1, we discuss the infimum of the empty set from the viewpoints of logic, common sense, and mathematics. In Section 2, we divide existence into three categories: *constructive existence*, *the existence derived from reduction to absurdity*, and *assumptive existence*. If possible, the existence should be constructed with the most effective algorithm. The solution of the hypergeometric equation can be represented by a series or an integral. Both representations belong to the constructive existence, but we would like to calculate the value of the hypergeometric function with the former representation because the former algorithm is more effective.

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.

If the derivation of an existence requires only a simple procedure, then it is called a close existence; if the derivation of an existence requires a complicated procedure, then it is called a distant existence, whether a constructive existence, an existence derived from reduction to absurdity, or an assumptive existence. We use the existence of prime factorization for cyclotomic integers as an example of constructive existence. In this example, we show how we preserve the *effectiveness* of divisibility tests for divisors when we divisors are generalized to ideals. After reviewing the entire scope of processes involving in constructing actual or ideal prime divisors under various conditions, we find that when we establish the most generalized theorem as our final result, we only preserve the least effective method and leave out the other more effective methods we have established along the way for specific cases. Thus, the climax of a theory is often a one-sided story or an oversimplification of a topic. Therefore, we should emphasize *methods* rather than a theory. The strength of a constructive existence depends on the tools we use to produce the existence. How we improve the *effectiveness* of a construction: bringing *trial and error* to a *level* as *basic* as possible, distinguishing what we can do from what we cannot, and eliminating unknown components in our solution wherever possible. The existence derived from reduction to absurdity originates from the elimination of impossibilities. The constructive existence is estimable, while the existence derived from reduction to absurdity is instimable. Let $\{\alpha_i \in \mathbb{R} | i \in I\} \neq \varnothing$. Then the existence of $\sup_{i\in I} \alpha_i$ is derived from reduction to absurdity [Rudin [36, p.11, l.−17–l.−16]]. We have no way to know its location on the real line. As we collect more $\alpha_i (i \in J \subset I)$ and find $\sup_{i\in J} \alpha_i$, this procedure will not help narrow down the search scope of $\sup_{i\in I} \alpha_i$. The proof of Fomin–Gelfand [12, p.198, Theorem] [the Ritz method] makes it easy to create computer simulations, while the proof of the statement given in Coddington–Levinson [3, p.197, l.7–l.8] does not. This is because the existence of each eigenvalue in the latter proof is derived from reduction to absurdity. The origins of assumptive existence: it is necessary to assume the existence of a mathematical object if we want to describe its general properties; the existence is provided by an axiom;

it is good for focusing on mathematical structures; it is an expedient choice for completeness of logic; we assume that a certain relation holds because it can be effectively proved for some special cases, but not for the general case (though the relation has no general proofs, after assuming the reation holds, we do have an interesting theory). The existence produced by the axiom of choice is assumptive. We should avoid using the axiom of choice whenever possible. In Section 3, we discuss generalization's influences on existence and uniqueness. In Section 4, we discuss how we effectively prove uniqueness, how we refine the proof of the uniqueness theorem, and why uniqueness can fail; then we explore possibilities for uniqueness-related problems. The use of the energy integral to prove the uniqueness of a mixed problem involving *the wave equation* provides strong support for the use of a system's *Hamiltonian* in establishing *the equation of motion* as a postulate in quantum mechanics. In Section 5, we conclude that, in order to make a theory effective, we should prioritize our choices in the following order: constructive existence, the existence derived from reduction to absurdity, and assumptive existence.

Keywords. Constructive existence, the existence derived from reduction to absurdity, assumptive existence, uniqueness, the Lebesque measure, essential supremum, mathematical induction, the axiom of choice, cyclotomic integers, prime factorization, ideal prime divisors, the periods of the cyclotomic field, prime ideals, divisibility tests, the level of trial and error, the strength of a constructive existence, Lyapunov's theorem, Dedekind's cuts, symmetric convergence, one-sided convergence, two-sided convergence, cut convergence, removable singularity, existence of a transcendental number, scope of the argument of reduction to absurdity, Zorn's lemma, maximal ideal, Chinese Remainder Theorem, Sophie Germain's theorem, Tauberian theorem, the Hahn–Banach theorem, Euclidean domain, principal ideal domain, Dedekind ring, refinement by weakening the hypothesis, initial value problems, the Lipschitz conditions, Cauchy's problem, Dirichlet's problem, growth conditions, boundary conditions, mixed problem for the wave equation, system's total energy, postulates in quantum mechanics, characteristic surfaces, successive approximations, Green's theorem, differential inequalities, the method of majorants, Fourier transforms, distribution solutions, the Ritz method, Sturm–Liouville problems, direct methods, method of finite differences, method of Lagrange multipliers, mean value property, methodical solutions

Before we discuss existence and uniqueness, let us study the infimum of the empty set from three different perspectives: logic, common sense, and mathematics. My point is that if we blindly perform mechanic operations in logic, we may easily overlook a natural approach and leave out its significant mathematical meanings.

1 The infimum of the empty set

Suppose $f : X \to [0, +\infty]$ is measurable. Let $\mathbb{R}$ be the real line and $S = \{\alpha \in \mathbb{R} | \, m(f^{-1}(\alpha, +\infty)) = 0\}$, where m denotes the Lebesque measure on $\mathbb{R}$. In order to define the concept of essential supremum of f (Rudin [38, p.67, l.16]), Rudin divides his discussion into two cases: $S \neq \varnothing$ and $S = \varnothing$. For the first case, he defines β to be $\inf S$. For the second case, in order to be consistent with the first case, he defines β to be $+\infty$ because every real number is a lower bound of S. The latter statement is derived from the following argument in logic: If there existed a real number γ which is not a lower bound of S, there would exist an element $\alpha \in S$ such that $\alpha < \gamma$. This contradicts $S = \varnothing$.

Next we consider the infimum of the empty set from the viewpoint of common sense. If $S = \varnothing, m(f^{-1}(\alpha, +\infty)) > 0$ for every $\alpha \in \mathbb{R}$. Namely, every $\alpha \in \mathbb{R}$ is not an essential upper bound of f. Since every $\alpha \in \mathbb{R}$ is too small to be an essential upper bound of f, we must choose the essential supremum β of f to be greater than or equal to every $\alpha \in \mathbb{R}$. Thus, β must be $+\infty$.

Finally, we consider the infimum of the empty set from the viewpoint of mathematics. In order to consider the essential supremum of f, by the definition of supremum (least upper bound), we must consider essential upper bounds of f. For $\alpha \in \mathbb{R} \cup \{+\infty\}$, if $f \leq \alpha$ a.e., α is called an essential upper bound of f. If α is a real number, then $f \leq \alpha$ a.e. if and only if $m(f^{-1}(\alpha, +\infty)) = 0$. Thus, $S \cup \{+\infty\} =$ the set of essential upper bounds of f.

Now let us consider the following two properties of essential upper bounds of real functions.

(a). $+\infty$ is an essential upper bound of every function:

(b). $+\infty$ is the only essential upper bound of f_0 , where $f_0 : [0, +\infty) \to [0, +\infty)$

$x \to [x]$, where $[x]$ is the greatest integer less or equal to x.

Rudin [38, p.67, Definition 3.7] is inadequate to vividly describe properties (a) or (b). Rudin reaches the definition of essential supremum without considering the complete set of essential upper bounds of f. Thus, his universal set is not inclusive. If Rudin were to naturally include $+\infty$ as an essential upper bound of f, he would not need to consider the infimum of the empty set.

2 The classification of existence

Before we discuss the classification of existence, let us list some *guidelines* that we should follow when studying existence:

1 We should narrow down the scope in which an existence is located. If A is a proper subset of B and the solutions of a problem belongs to A, then we say the solutions are in A rather than B.

2 We should compare the tools we use to produce various types of existence. This analysis can be used to determine the relative strength of an existence.

3 We should avoid using an *infinite process* such as the mathematical induction or the axiom of choice wherever possible.

Example 2.1. In Rudin [36, p.34, l.-5], Rudin uses mathematical induction to construct the sequence $\{I_n\}$ in order to prove that for every open cover of I there exists a finite subcover (In Rudin [36, p.34, Theorem 2.40]). In contrast, Johnson proves Johnson–Kiokemeister–Wolk [25, p.147, Theorem 6.3] without using mathematical induction.

In order to reduce confusion and misunderstanding, we should at least divide existence into the following three categories: constructive existence, the existence derived from reduction to absurdity, and assumptive existence.

2.1 Constructive existence

Let Statement $A = x$ has property P

and Statement $B = x$ does not have property P.

We should be able to determine whether each of the two statements A and B is either true or false. If and only if A is true, then B is false.

We may use various methods to effectively construct an element x which has property P.

2.1.1 Effectiveness of a construction

2.1.1.1 Constructions vs. resources

Mathematical resources are richer than logical resources. Physical resources are richer than mathematical resources. The constructive existence becomes more effective if we are allowed to use a richer set of resources. For example, we use an electric charge to test whether an electric field exists, and then measure its size. This approach is consistent with Section 2, Guideline 1.

Karamcheti [27, p.166, l.−5–l.−1] is an awkward way to say that the existence of a dynamic solution is more effective than the existence of a kinematic solution.

2.1.1.2 If possible, the existence should be constructed with the most effective algorithm

The solution of the hypergeometric equation can be represented by a series or an integral. Both representations belong to the constructive existence, but we would like to calculate the value of the hypergeometric function with the former representation because the former algorithm is more effective: Let $u(z) = \int_0^1 F(z,t)dt = \sum_{n=0}^{\infty} a_n(z,t)(z-z_0)^n$. Suppose we want to calculate $u(z)$ in a neighborhood of $z = z_0$. If we choose the integral representation, we have to approach the integral with the Riemann sum for each z_k. Thus, we will repeat the infinite process many times. In contrast, if we choose the series representation, the infinite process has to be done only once because the series applies to the entire neighborhood. This is the reason why we use Stirling's formula [Watson–Whittaker [45, p.279, l.4]] instead of the definition given in Watson–Whittaker [45, p.243, l.3] to calculate the gamma function.

Lebedev [29, §9.5] shows how to represent the hypergeometric function as a series in the domain $T\{z : |z| < 1\}$, where T is a linear transformation given in Lebedev [29, p.246, l.−14]. Lebedev [29, §9.7] uses L'Hospital's rule and Lebedev [29, p.3, (1.2.2)] to discuss the exceptional cases left out by Lebedev [29, §9.5].

Remark 1. Watson–Whittaker [45, p.288, l.3] gives an explicit formula for the analytic continuation of the hypergeometric function. Lebedev [29, p.240, l.−18–p.241, l.6] gives a more elementary method of carrying out the analytic continuation. If we reguard calculating the values of the hypergeometric function as our purpose, then the latter method is more effective. This is because it suffices to represent the integral as series in a more restrictive domain $\Re\gamma > \Re\beta > 0, |\arg(1-z)| < \pi$.

Remark 2. Lebedev [29, p.256, (9.7.3)] is proved as follows:

Proof. Case $0 \le k < n$: $\frac{\partial}{\partial\gamma}[\Gamma(1+\alpha+\beta-\gamma+k)]^{-1}|_{\gamma=\alpha+\beta+n} = (-1)^{n-k}(n-k-1)!$ [Lebedev [29, p.3, (1.2.2)]].

Case $k \ge n$: $\frac{\partial}{\partial\gamma}[\Gamma(1+\alpha+\beta-\gamma+k)]^{-1}|_{\gamma=\alpha+\beta+n} = \frac{\psi(1-n+k)}{\Gamma(1-n+k)}$. □

2.1.1.3 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 [45, p.339, l.−13–l.−12]] follows from Watson–Whittaker [45, p.292, l.−15–l.−10] and Guo–Wang [15, p.95, l.−8].

Remark. (Methodical solutions) The differential equation given in Watson–Whittaker [45, p.291, l.-11–l.-7] belongs to a special type. The given solution is justified simply by substitution [Watson–Whittaker [45, 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 [45, p.339, l.-13–l.-12] cannot be considered a methodical solution. In contrast, the integral solution given in Guo–Wang [15, 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 [15, §2.13]]. In fact, the integrand and the path of integration [Guo–Wang [15, p.302, l.4–l.13]] can be specified by the Laplace transform. Consequently, the latter solution is more methodical than the former one.

2.1.1.4 The developing process of a construction

By following two steps below, we establish the existence of prime factorization for cyclotomic integers.

2.1.1.4.1 Step 1: Factoring ordinary prime integers into ideal prime divisors

Let λ be an ordinary prime integer and let α satisfy $\alpha^{\lambda} = 1$. Given the ring of cyclotomic integers (Edwards [10, p.81, l.-8]), we try to find prime cyclotomic integers (Edwards [10, p.84, l.24]).

Problem 1. From where should we begin our task of factorization?

Solution. Since the ordinary primes are the basic units of prime factorization for ordinary integers, we try to break down an ordinary prime p further into prime divisors of p (i.e. prime cyclotomic integers which divide p; we call them actual prime divisors of p). We assume they exist and then use analysis to find necessary conditions: the theorem given in Edwards [10, p.90, l.-3–p.91, l.4] and the theorem given in Edwards [10, p.92, l.-12–l.-7]. Based on these necessary conditions, we may use synthesis to establish an effective method of finding a prime cyclotomic integer (Edwards [10, p.92, l.-3–l.-1]) whose norm is p for the case $p \equiv 1 \bmod \lambda$. Using the theorem given in Edwards [10, p.92, l.-3–l.-1] Lamé found some prime cyclotomic integers as shown in Edwards [10, p.99, Table 4.4.1]. □

Problem 2. For the case $\lambda = 7$, we cannot find a factorization of 29 by Lamé's method (Edwards [10, p.99, l.-20–l.-13]). What technique can we use to factor 29?

Solution. Kummer developed a technique (Edwards [10, p.99, l.-9]) to find some k's which satisfy the condition given in Edwards [10, p.91, l.2]. Then he succeeded in obtaining $\alpha^2 - \alpha^4 + 1$ as a prime divisor of 29. Kummer noted that in the case $\lambda = 23$, 47 has no prime divisor which is a cyclotomic integer (Edwards [10, p.105, l.23]) and that the product $47 \cdot 139$ has two different ways of factoring into irreducible cyclotomic integers (Edwards [10, p.105, l.-1–p.106, l.4]). In case $\lambda = 13, \gamma = 2$, and $e = 4$, the task of computing the norm can be reduced to three lines (Edwards [10, p.104, l.10–l.12]) if we use the periods of length f $(= 3)$ (Edwards [10, p.108, l.-11]). The periods of length f are the cyclotomic integers invariant under σ^e (Edwards [10, p.108, l.17–l.19]). □

Remark. Now let us read van der Waerden [42, vol. 1, §54]. We find that van der Waeden fails to show the role periods play in finding prime cyclotomic integers.

Problem 3. We can find factorizations for p when $p = \lambda$ or $p \equiv 1 \bmod \lambda$. How do we find factorizations for p when $p \not\equiv 1 \bmod \lambda$?

Solution. Suppose p is divisible by a hypothetical prime cyclotomic integer $h(\alpha)$ and is not congruent to 1 mod λ. Then ($g(\alpha)$ is congruent to an integer u *mod* $h(\alpha)$) $\Rightarrow$ ($g(\alpha)$ is made up of periods of length f) (Edwards [10, p.112, l.1–l.2; l.10–l.11]). Kummer proved that the inverse of the previous statement is also true (Edwards [10, p.112, l.−13–l.−12]). This shows that the concept of periods plays an even more important role in the case that $p \not\equiv 1 \bmod \lambda$. Now we may use analysis to give strong necessary conditions satisfied by a prime cyclotomic integer as shown in Edwards [10, p.113, l.18–l.27]. Then our remaining task is to prove the synthesized statement given in Edwards [10, p.113, l.−11–l.−9]. For $\lambda = 5$ and $p = 29$, the trial-and-error method of building the table (Edwards [10, p.115, Table 4.7.1]) does not help much. It would be better to find the integers u_0 and u_1 given in Edwards [10, p.112, Theorem]. The obstacle is overcome by reducing e congruences in e unknowns to one congruence in one unknown (Edwards [10, p.115, l.−11; p.118, l.1–l.10]). □

Remark. In the course of developing a theory, we should look for patterns. The congruence relations
$\eta_i \equiv u_i \bmod h(\alpha)$ in Edwards [10, p.112, l.−14] are similar to the congruence relation $\alpha \equiv k \bmod h(\alpha)$ in Edwards [10, p.90, l.−1–p.91, l.1]. See Solution to Problem 2.

Problem 4. For the case $\lambda = 31$ and $p = 2$, we can only prove that 2 has no factorization whose factors are made up of periods of length 5. This is because we only have the divisibility test for periods (Edwards [10, p.112, l.−16–l.−12; p.113, l.−11–l.−9]). How do we extend the test to all the cyclotomic integers?

Solution. Kummer overcame this obstacle by establishing the theorem and its corollary given in Edwards [10, p.123, l.−20–l.−9]. □

Problem 5. In order to apply the divisibility test given in Edwards [10, p.123, Corollary], we must find the integers $u_1, u_2, \cdots, u_e$. However, the existences of the integers $u_1, u_2, \cdots, u_e$ require the assumption that p has a prime divisor which is a cyclotomic integer. In Solution to Problem 2 we see that in the case $\lambda = 23$, 47 has no prime divisor which is a cyclotomic integer. How do we extend the divisibility test to the case where no cyclotomic integer is a prime divisor of p?

Solution. First, we construct $\psi(\eta)$ and u_i $(i = 1, \cdots, e)$ so that $\eta_i \psi(\eta) \equiv u_i \psi(\eta) \bmod p$ (Edwards [10, p.129, l.22]). Second, we define the congruence modulo the prime divisor of p corresponding to $u_1, u_2, \cdots, u_e$ as shown in Edwards [10, p.130, l.1–l.2]. This congruence relation is prime (Edwards [10, p.127, l.2]). There are precisely e prime divisors of p ; a cyclotomic integer $g(\alpha)$ is divisible by p if and only if $g(\alpha)$ is divisible by all e prime divisors of p (Edwards [10, p.128, Theorem 2]). □

Remark. We call a prime divisor defined in Edwards [10, pp.136–137, Definition] an ideal prime divisor. The problem of factoring cyclotomic integers into ideal prime divisors is essentially solved. Edwards [10, p.128, l.9–l.11; p.135, l.-11–l.-10; p.136, l.-4] provide the existence of ideal factorization. Edwards [10, p.138, Corollary] provides the uniqueness of the factorization. In Solution to Problem 2 we note that in the case $\lambda = 23$, 47 has no actual prime divisors. Now 47 can be factored into 22 ideal prime divisors. So can 139. None of these 44 ideal prime divisor is an actual prime divisor. $1-\alpha+\alpha^{21}$ can be expressed as the product of one ideal prime factor of 47 and one ideal prime factor of 139 (Edwards [10, p.105, l.12; l.-19–l.-15]). In terms of the notations given in Edwards [10, p.141, l.-12], we have
$1-\alpha+\alpha^{21} = (47, 1-\alpha+\alpha^{21})(139, 1-\alpha+\alpha^{21})$. This explains why there are two different ways to factor $47 \cdot 139$ into irreducible cyclotomic integers.

Problem 6. How do we express ideal prime divisors of p in terms of cyclotomic integers?

Solution. An arbitrary divisor can be written in the form given in Edwards [10, p.141, l.-12]. □

Discussion:

(1) Ensuring that we have not broken down factors more than necessary

Ordinary prime integers play the role of molecules in factoring integers into prime factors. Factoring cyclotomic integers into ideal prime divisors is like further breaking molecules into atoms. After the breakdown, we must make sure that the ideal factorization is consistent with the actual factorization and that the breakdown is not extensive than necessary: if a cyclotomic integer $g(\alpha)$ is divisible by another cyclotomic integer $h(\alpha)$ in the sense of ideal divisors (Edwards [10, p.139, l.22]), then there exists a cyclotomic integer $k(\alpha)$ such that $g(\alpha) = h(\alpha)k(\alpha)$. That is, we must make sure that the quotient $k(\alpha)$ is a *whole* rather than a *fractional* cyclotomic integer. This is the content of Edwards [10, p.137, the fundamental theorem]. See Edwards [10, p.139, l.-13–l.-12]. The verification can be done by clarifying the meanings of the following statements: When Edwards says that a cyclotomic integer $g(\alpha)$ is divisible by p in the ordinary sense (Edwards [10, p.128, l.9–l.10]), he means that there exists a cyclotomic integer $h(\alpha)$ such that $g(\alpha) = p\, h(\alpha)$. This also clarifies the meaning of the statement that $h(\alpha)$ is divisible by p in Edwards [10, p.138, l.2–l.3]. Therefore, when Edwards says that $g(\alpha)$ divides $h(\alpha)$ in Edwards [10, p.139, l.-13], he means that there exists a cyclotomic integer $k(\alpha)$ such that $h(\alpha) = k(\alpha)g(\alpha)$.

(2) Solving a more complicated problem requires more effort and more refined strategies

Like overcoming an obstacle, each problem requires some effort and new strategies to find its solution. Solution to Problem $(i+1)$ requires more effort than Solution to Problem i, so Solution to Problem $(i+1)$ is more sophisticated and less effective than Solution to Problem i. Solution to Problem $(i+1)$ can solve problems that Solution to Problem i cannot, so the method of solving Problem $(i+1)$ is more refined than the method of solving Problem i. The congruence relation modulo an actual prime divisor of p implies (Edwards [10, p.126, l.11–l.12]) one of the e prime congruence relations given in Edwards [10, p.128, l.7], so the prime divisor of p corresponding to $u_1, u_2, \cdots, u_e$ is *weaker* and *more generalized* than the actual prime divisor of p. If both Solution to Problem i and Solution to Problem j, where $j > i$, can be used to solve Problem i, we should use Solution to Problem i rather than Solution to Problem j, because the latter is less effective for Problem i.

2.1.1.4.2 Step 2: Factoring cyclotomic integers into prime ideals

In order to apply the above methods to other types of algebraic integers (Edwards [10, p.144, l.−3]), we should generalize the concept of ideal prime divisors to that of prime ideals.

(1) We try to identify divisors (Edwards [10, p.139, l.14–l.15]) with ideals (Edwards [10, p.144, l.19–l.−8]).

In order to define division among ideals, we first characterize the concept among principal ideals: $p|q$ if and only if $\langle q\rangle \subseteq \langle p\rangle$. The first method: it *seems* natural and *harmless* to define $A|B$ if and only if $B \subseteq A$ (Stewart–Tall [41, p.121, Proposition 5.6]). However, if we were to adopt this approach, we would actually end up *losing* all the *effective* divisibility tests for cyclotomic integers. This is because we have no effective algorithm to check whether every element of B is contained in A. In order to preserve the effectiveness of divisibility tests, it is unnecessary to consider all the ideals. It suffices to consider all the prime ideals. The second method: all we need to do is identify ideal prime divisors with prime ideals and then follow the scheme given in Edwards [10, §4.12] to define division for ideals. The second method allows us to *preserve* all the effective divisibility tests for divisors when we replace divisors with ideals.

Remark. For cyclotomic integers, we have the following effective divisibility tests for divisors: Edwards [10, p.90, Theorem] (for actual prime divisors if $p \equiv 1 \bmod \lambda$) $\longrightarrow$ Edwards [10, p.121, l.−12–l.−8; p.123, Theorem and its Corollary] (for actual prime divisors if $p \not\equiv 1 \bmod \lambda$) $\longrightarrow$ Edwards [10, p.126, Proposition] (for ideal prime divisors) $\longrightarrow$ Edwards [10, p.139, l.22] (for ideal divisors).

(2) We try to translate the ideal prime factorization given in 2.1.1.3.1 in the language of modern field theory.

Based on the way that $P(X)$ is constructed in Edwards [10, p.122, l.5], the irreducible polynomial $X^{p-1}+X^{p-2}+\cdots+X+1$ in $\mathbb{Q}[X]$ can be expressed as the product of $P(X)$ [e.g. $\alpha^3-\eta_0\alpha^2+\eta_2\alpha-1$ (Edwards [10, p.122, l.10])] and its conjugates in $\mathbb{Q}[\eta_0][X]$. $P(X)$ is the irreducible (van der Waerden [42, vol.1, p.165, l.−13; p.97; l.−6–l.−5]) polynomial over $\mathbb{Q}[\eta_0]$ [a normal extension over $\mathbb{Q}$ (van der Waerden [42, vol.1, p.164, l.6; p.160, l.1–l.2])] of degree f with α as a root (Edwards [10, p.122, l.5]). By Edwards [10, p.127, l.4; p.146, Theorem] and Stewart–Tall [41, p.187, l.14–l.15], $\bar{P}(X)$ (e.g. $X^3-u_0X^2+u_2X-1$) is irreducible $\bmod p$, where $\bar{P}(X)$ (Stewart–Tall [41, p.186, l.8]) is obtained from $P(X)$ by replacing its coefficients of the form $g(\eta)$ with the coefficients of the form $g(u)$. We could give a more structured and sophisticated explanation using Galois theory, but it may distract us from the point under discussion and obscure its essence.

Remark 1. The construction of $\varphi(X)$ given in van der Waerden [42, vol.1, chap. V, §32] and the factorization of $\bar{f}$ given in Stewart–Tall [41, p.187, l.−3–p.188, l.3] are ineffective. In contrast, the acquisitions of $\alpha^3-\eta_0\alpha^2+\eta_2\alpha-1$ (Edwards [10, p.122, l.−13–l.−12]) and u_i (Edwards [10, p.120, l.1–l.7]) are effective.

Remark 2. For the case of cyclotomic integers, the $e_1,\cdots,e_r$ given in Stewart–Tall [41, p.186, l.8] all equal 1 (van der Waerden [42, vol.1, p.120, l.4–l.19 or p.124, l.14–l.15]).

Discussion: A theorem is often an *oversimplification* in summarizing various methods under discussion

Constructing a prime ideal containing $\langle p\rangle$ is more sophisticated and less effective than constructing an actual prime divisor of p.

Modern algebra textbooks often emphasizes Kummer's achievements by listing the theorems he established. In fact, the concept of the unique factorization theorem can be easily acquired from natural numbers rather than from cyclotomic integers. In addition, Kummer's method can produce the strong form of factorization (i.e. a product of prime cyclotomic integers) in some cases (Edwards [10, p.116, l.13]), while his unique factorization theorem (Stewart–Tall [41, p.186, Theorem 10.1; p.117, Theorem 5.5]) can merely produce the weak form of factorization (i.e. a product of prime ideals). In fact, even for the simple case $N(\alpha-1)=5$ (Edwards [10, p.92, l.−10–l.−9; p.93, l.1]), Stewart–Tall [41, p.186, Theorem 10.1] is inadequate for producing the strong form of factorization. Consequently, in studying Kummer's theory what we should focus on are the obstacles he encountered and the methods he used to solve his problems. A theory guided by questions and answers is naturally organized. To build a theory through any other means only makes it artificial and disorganized.

A mathematical theory is a side product of the methods it uses. In some specific cases, we may have more effective methods. In order to cover all the cases, Stewart–Tall [41, p.186, Theorem 10.1] can only present the most ineffective method. Thus, Stewart–Tall [41, p.186, Theorem 10.1] is an oversimplification in summarizing the various methods given in Edwards [10, chap. 4]. Applying a theory to a practical problem is usually not as effectively as applying the method used to develop the theory to the problem. Thus, in mathematics we must emphasize *methods* instead of theories.

In the theory of partial differential equations, we encounter a similar situation: Sneddon [40, p.50, Theorem 2] $\longrightarrow$ Petrovsky [33, pp.15–16, the Cauchy–Kowalewski Theorem] $\longrightarrow$ Rudin [37, p.195, Theorem 8.5 (the existence of a fundamental solution)]. The first theorem provides a solution in closed form. The second theorem uses the method of majorants to acquire a solution in power series. The third theorem uses Fourier transforms to find a distribution solution. If we try to solve a wider class of partial differential equations, the method we use will become less effective. Note that the operation of differentiation of a distribution is weaker than the operation of differentiation of a continuously differentiable function (Rudin [37, p.136, l.13–l.26]). Note also that the axiom of choice (Rudin [37, p.58, l.3; p.57, l.11–l.13]) is used in Rudin [37, p.197, l.13]. If we just discuss classical solutions or distribution solutions without including all the methods, the discussion is incomplete.

2.1.1.5 How we improve the effectiveness of a construction

2.1.1.5.1 Bringing trial and error to a level as basic as possible

When we construct a mathematical element in a finite number of steps, we should bring trial and error to a level as basic as possible. This will help trace its origin, reveal inner structures, and motivate us to construct the element in a more specific and effective manner.

Example 2.2. We construct a prime divisor (Edwards [10, p.106, l.−8–l.−7]) in the ring of cyclotomic integers (Edwards [10, p.81, l.−8]) through the following steps:

Level 1. Defining prime divisors: use Edwards [10, p. 126, Proposition] to find a set of integers $u_1, u_2, \cdots, u_e$.

Level 2. Factoring the polynomial $X^{\lambda-1}+X^{\lambda-2}+\cdots+X+1$ into irreducible polynomials $\bmod p$ by Edwards [10, p. 125, Exercise 3]. See 2.1.1.3.2, (2).

Level 3. Using Stewart–Tall [41, p.186, Theorem 10.1] to find prime ideals of $\langle p\rangle$. See 2.1.1.3.2, Remarks.

For Level 1, if p has a prime divisor which is a cyclotomic integer, we have a more effective method (Edwards [10, p.126, l.4–l.5]) of finding $u_1, u_2, \cdots, u_e$ than the trial-and-error method given in Edwards [10, p.126, Proposition]. See Edwards [10, p.118, l.17; p.120, l.5]. Stewart omits Level 1, and use the trial-and-error method in Level 2 (Stewart–Tall [41, p.187, l.−3–p.188, l.3]). Thus, the features (Edwards [10, p. 125, Exercise 3]) of the periods of a cyclotomic field given in van der Waerden [42, vol.1, p.165, l.-8] is never revealed in Stewart's presentation. This omission somehow eliminates the motive to create a prime divisor from the feature of an actual prime divisor of p. See Edwards [10, p.127, l.12–l.14].

2.1.1.5.2 We should be able to distinguish what we can do from what we cannot

Example 2.3. (Maximal solutions)(Hartman [18, p.25, Lemma 2.1])

The definition Φ given in Coddington–Levinson [3, p.46, l.5] is problematic because there exists an initial value problem that we know only some of its solutions, but not all of them. In such a case, we may find a supremum of some solutions, but we cannot find the supremum of all solutions. Someone may argue that the shortcoming is amended by proving $\Phi = \lim_{m\to\infty} \varphi_{1/m}$ (Coddington–Levinson [3, p.47, l.2–l.3]). However, $\varphi_{1/m}$ is not accessible because φ_ε is defined by means of φ_i $(i = 0, 1, \ldots, n-1)$ and φ_i's depend on Φ which is not accessible. The construction given in Hartman [18, p.25, (2.4)] enables us to avoid considering all the solutions.

2.1.1.5.3 We should eliminate unknown components in our solution wherever possible

An effective construction takes full advantage of available resources, while generalization often reduces a theorem's assumption to the minimum.

Theorem 2.1. (Lyapunov's theorem)(Coddington–Levinson [3, p.314, l.-10–p.315, l.21])

Remark 1. The solution of the problem given in Coddington–Levinson [3, p.315, l.16] can be found in Wikipedia [47, §Integral form for continuous functions, (b)].

Remark 2. Let σ in Coddington–Levinson [3, p.315, (1.4)] be the μ in Coddington–Levinson [3, p.315, (1.5)]. See Coddington–Levinson [3, p.316, l.3–l.4]. In contrast, we cannot effectively construct α in Pontryagin [35, p.211, l.-11]. If we trace its construction (Pontryagin [35, p.210, l.13; p.208, l.2; p.206, l.4–l.5]), we will find that its existence is derived from reduction to absurdity (Rudin [36, p.31, l.12–l.16]). The proof pattern of Pontryagin [35, p.208, Theorem 19] is the same as that of Hartman [18, p.40, Theorem 8.4]. The latter theorem is more generalized, so the derived constant is less effective. There is another drawback in the proof of Pontryagin [35, p.208, Theorem 19]: W in Pontryagin [35, p.207, (20)] is determined by ψ_i, and ψ_i is determined by Pontryagin [35, p.206, (16)]. In other words, μ and ν in Pontryagin [35, p.206, (13)], and thereby the α in Pontryagin [35, p.210, l.14] are all determined by Pontryagin [35, p.206, (16)]. Consequently, if we want to determined the α in Pontryagin [35, p.210, l.14], we must slove the differential equation: Pontryagin [35, p.206, (16)]. However, our goal is to determine the solution's stability without actually solving the differential equation.

2.1.1.5.4 We should not add any more structures than necessary as we proceed toward our goal

Example 2.4. (Effective constructions)

The goal of a construction determines what method we should use. We should not add any more structures than necessary as we proceed toward our goal. This approach allows us to easily recognize the role of each piece (e.g., the agreement of Landau [28, p.78, Definition 54] with Landau [28, p.53, Definition

35]) of the construction and helps us gain insight into the structures of irrational numbers. For example, when we try to extend the set of positive rational numbers to the set of positive real numbers, our goal is to ensure the validity of following statement: Suppose S is a subset of positive real numbers; then there exists a least upper bound of S if S is bounded above.

Thus, the above problem is essentially a problem concerning ordering or positiveness. Therefore, the tool of Dedekind's cuts (Landau [28, p.43, Definition 28]) is a more appropriate choice for us to construct irrational numbers than the tool of Cauchy sequences (van der Waerden [42, vol.1, p.212, l.−18]). The tool of Cauchy sequences is designed for the completion of a metric space rather than the construction of irrational numbers. In other words, it is designed for topological extensions rather than algebraic extensions. In fact, only after we detach unnecessary complications such as negative numbers, topology, metric spaces, and infinitesimals may we firmly grasp the essence of positive irrational numbers (Landau [28, p.67, Definition 42]). In view of Landau [28, p.9, Definition 2], it is incorrect to say that ordering is a non-algebraic concept (van der Waerden [42, vol.1, p.209, l.2]).

2.1.1.6 The strength of a constructive existence

The strength of a constructive existence depends on the tools we use to produce the existence. If the convergence of Type I implies the convergence of Type II, then we say the convergence of Type I is stronger than the convergence of Type II. When we say an improper integral *exists*, we mean that the integral *converges* to a certain limit. The strength of the existence of the integral value depends on the strength of the convergence we adopt as our *method* to evaluate the integral. When evaluating an integral, we should not only calculate its value, but also point out the integral type which allows the *strongest convergence* to that value.

(1) The absolute convergence is stronger than the ordinary convergence.

The absolute convergence of the integral given in Ahlfors [1, p.154, Example 2] can be proved using the condition given in González [14, p.689, l.12].

(2) The asymmetric (ordinary) convergence is stronger than the symmetric convergence (González [14, p.686, l.1]).

The use of a semicircle can only prove the symmetric convergence of the integral given in González [14, p.693, (9.11-11)], while the use of a rectangle proves the asymmetric convergence of the integral given in González [14, p.695, (9.11-15)]. Ahlfors [1, p.156, Fig. 25] is used to prove the asymmetric convergence of the limit given in Ahlfors [1, p.157, l.4], while González [14, p.698, Fig. 9.16] is used to prove the symmetric convergence of the integral given in González [14, p.698, (9.11-21)].

(3) The one-sided convergence ($\lim_{R\to\infty}\int_{[0,R]}$) is stronger than the two-sided convergence ($\lim_{\varepsilon\to0+,R\to\infty}\int_{[\varepsilon,R]}$).

Ahlfors finds the value of $\lim_{R\to\infty}\int_{[0,R]}(\sin x)x^{-1}dx$ in Ahlfors [1, p.157, l.−11] using the concept of removable singularity. In contrast, González only finds the value of $\lim_{\varepsilon\to0+,R\to\infty}\int_{[\varepsilon,R]}(\sin x)x^{-1}dx$ in González [14, p.703, (9.11-28)].

(4) The non-cut convergence is stronger than the cut convergence (González [14, p.686, 2.(c)]).

Ahlfors finds the value of $\lim_{M\to-\infty,N\to\infty}\int_{[M,N]}(\sin x)x^{-1}dx$ in Ahlfors [1, p.157, l.15], while González only finds the value of $\lim_{\varepsilon\to0+,R\to\infty}(\int_{[-R,-\varepsilon]}+\int_{[\varepsilon,R]})(\sin x)x^{-1}dx$ in González [14, p.703, (9.11-27)].

Remark. (Slow convergences vs. fast convergences) If two sequences converge to the same number and the second sequence can attain the prescribed accuracy with fewer terms, then we say the second sequence converges faster than the first one. Example: Guo–Wang [15, p. 13, l.1–p. 13, l.−5].

2.2 The existence derived from reduction to absurdity

There are two ways to solve a problem: First, we may explore the possibilities such as constructing the solution using a certain resource; second, we may eliminate the impossibilities using reduction to absurdity. The latter option may proceed as follows: if we assume that $\{x \,|x \text{ has property } P\} = \varnothing$, we will have a contradiction. Thus, this existence of x with property P is a purely logical consequence. The negation of the statement "every x does not have property P" makes the existence of x indefinite (in terms of the Venn diagram, we can find the scope it belongs, but we cannot locate its position) and fails to provide an effective method to find a particular element with property P. For a given object, sometimes we do not have enough information to determine whether or not it has property P. Therefore, an existence derived from reduction to absurdity is weaker than a constructive existence.

Example 2.5.
There are three people and two seats in a room. Therefore, there is one person who has no seat.

Example 2.6. (The existence of a transcendental number proved by Cantor's argument) (Hardy–Wright [17, p.160, Theorem 190])

Example 2.7.
Given Coddington–Levinson [3, p.194, l.1–l.2], we want to find the essential differences between Ince [20, p.233, Theorem II] and the statement given in Coddington–Levinson [3, p.197, l.7–l.8] and the reason that the former proof is better.

(1) Former proof:

  (a) Prerequisite: Ince [20, p.229, l.16–l.25]
  (b) Find μ_0 (Ince [20, p.233, l.15]), and then λ_0 (Ince [20, p.233, l.26–l.27]).
  (c) Given μ_m (Ince [20, p.232, l.1–l.13]), we use Ince [20, p.232, l.14–l.33] to find $\lambda_{m+1} \in (\mu_m, \mu_{m+1})$.
  (d) The existence of eigenvalues is intuitive and essentially constructive; the proof also provides the information about the number of zeros.

(2) Latter proof:

  (a) In order to construct the Green function G (Coddington–Levinson [3, p.193, l.11]), and thereby, $\mathscr{G}$ (Coddington–Levinson [3, p.193, l.15]), we have to assume $l = 0$ is not an eigenvalue (Coddington–Levinson [3, p.193, l.4]). By Coddington–Levinson [3, p.189, Theorem 2.1], this can be done (Coddington–Levinson [3, p.193, l.5–l.9]). However, the existence of a non-eigenvalue is derived from reduction to absurdity.
  (b) By Coddington–Levinson [3, p.196, l.14], we may consider eigenvalues of $\mathscr{G}$ in $(0, ||\mathscr{G}||)$ without loss of generality.
  (c) The proof of Rudin [36, p.11, Theorem 1.36] uses reduction to absurdity to prove the existence of the supremum. From the definition of supremum, we see the existence of $u_m \in C$ given in Coddington–Levinson [3, p.195, l.−12] is derived from reduction to absurdity.

(d) By Coddington–Levinson [3, p.196, l.−7], $|\mu_m|$ decreases to zero. By Coddington–Levinson [3, p.194, l.1–l.2], this agrees with Coddington–Levinson [3, p.189, Theorem 2.1].

(e) The existence of eigenvalues is not constructive. The proof does not provide the information about the number of zeros.

For the existence derived from reduction to absurdity, we should restrict the scope of the argument of reduction to absurdity to make it as small as possible.

Example 2.8.
The argument of Niven–Zuckerman [31, p.144, Theorem 5.6] is better than that of Hua [19, p.208, Theorem 7.8] because Niven reduces the scope of the argument of the reduction to absurdity from the set of natural numbers to the set of prime numbers (Niven–Zuckerman [31, p.143, Theorem 5.5; p.142, Theorem 5.3]).

Reducing the scope of the argument of reduction to absurdity may help us locate the solutions more precisely.

Example 2.9.
The proof of Niven–Zuckerman [31, p.224, Theorem 8.6] is better than that of Hua [19, p.76, Theorem 4.3] because the former proof reduces the indefinite set from $(n,+\infty)$ to $(n,2n)$, where n is a positive integer greater than 1.

Example 2.10. (Estimable vs. Inestimable)
The existence of M given in González [13, p.639, l.−5] is derived from reduction to absurdity. See the proofs of Rudin [36, p.11, Theorem 1.36; p.31, Theorem 2.28; p.35, Theorem 2.41]. Consequently, the M is instimable. In constrast, the existence of M given in Guo–Wang [15, p.151, l.2–l.3] is constructive. Therefore, the M is estimable.

Example 2.11. (The Ritz method is an effective tool for studying Sturm–Liouville Problems [Fomin–Gelfand [12, pp.198–205, §41]])
I. Calculus tools for finding extrema of functions: Kaplan [26, §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 [7, vol.1, chap. IV, §2]].
II. Solving Sturm–Liouville Problems effectively [Fomin–Gelfand [12, pp.196–197, Remark 2]] by the Ritz method [Fomin–Gelfand [12, p.196, Theorem]]: construct a complete sequence of functions φ_n as in Fomin–Gelfand [12, 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 [12, p.195, (10)]]. Thus, it suffices to calculate y_n given in Fomin–Gelfand [12, 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 [12, p.200, (24)] is more constructive and effective than the existence of μ_0 given in Coddington–Levinson [3, p.195, l.−9].

Explanation. (A).
1. M defined as in Fomin–Gelfand [12, 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 [12, 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

[26, §2.19; §2.20].
3. Define $\lambda_n^{(1)}, y_n^{(1)} \quad (n = 1,2,\cdots)$ as in [Fomin–Gelfand [12, p.199, l.−10–l.−7]]. Then $\lambda_{n+1}^{(1)} \leq \lambda_n^{(1)}$ [Fomin–Gelfand [12, p.200, (23)]]. Define $\lambda^{(1)}$ as in Fomin–Gelfand [12, 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 [12, p.201, l.−14–p.203, l.−3], we use the method of Lagrange multipliers to obtain Fomin–Gelfand [12, p.203, (36)] and then use Fomin–Gelfand [12, p.201, Lemma 2] to prove Fomin–Gelfand [12, 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 [3, p.195, l.2; l.−9]]. The existence of supremum is derived from reduction to absurdity [Rudin [36, 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 [3, 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 [12, §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 [3, 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 [3, 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. □

2.3 Assumptive existence

Assumptive existence is often used in the following situations:

(1) It is necessary to assume the existence of a mathematical object if we want to describe its general properties.

Example 2.12. (The limit of the sum equals the sum of limits)
In Rudin [36, p.43, Theorem 3.3(a)] are assumptive, the existence of $\lim_{n\to\infty}(s_n + t_n)$ is a consequence of the assumption that s and t exist.

(2) An assumptive existence is provided by a widely accepted axiom which we believe is true.
In this case, the existence is assumed by an axiom without proof. In other words, we assign the logical value of the axiom to be true by assumption. For example, many mathematicians assume that the axiom of choice is true. This approach allows us to expand our research to the subject that requires the axiom of choice. The picture may become somewhat vague, but with these new results we may achieve some artistic purpose such as extension, unification, and completion.
The proof of McCoy [30, p.458, Lemma 17.38] uses Zorn's lemma which is equivalent to the axiom of choice (Dugundji [9, pp.31–32, Theorem 2.1]). In my opinion, although Stewart–Tall [41, p.86, Proposition 4.5 (b) ⇒ (c)] is proved without using the axiom of choice, the proof provides less information than that of McCoy [30, p.458, Lemma 17.38] about how we construct a maximal ideal. Thus, an assumptive existence is not necessary worse than an existence derived from reduction to absurdity.

(3) When the main point of a mathenatical theory is about its structure rather than its methods's effectiveness, we use assumptive existences.

Example 2.13. (The existence of a topology) (Pervin [32, p.36, l.−2–p.37, l.5])

(4) We use an assumptive existence because it only appears in a transitory stage.

Sometimes an assumptive existence is an expedient choice to avoid trouble. In proving Ellison–Ellison [11, p.277, Theorem 8.8(b)], we use Ellison–Ellison [11, p.258, Theorem 8.4] (See Ellison–Ellison [11, p.277, l.16]). The proof of Ellison–Ellison [11, p.258, Theorem 8.4(a)] is divided into two cases: in case (i) (Ellison–Ellison [11, p.258, l.−2]), $c(\varepsilon)$ exists; in case (ii) (Ellison–Ellison [11, p.258, l.−1]), $c(\varepsilon, k_0, \beta_0)$ exists.

Because we do not know whether there exists a β (Ellison–Ellison [11, p.258, l.−9]), we may assume it does for the completeness of logic. Even though the assumption that a β exists will not affect the validity of Ellison–Ellison [11, p.277, Theorem 8.8(b)], it does affect the effectiveness of A_1 (Ellison–Ellison [11, p.278, l.7]) and $c(\varepsilon)$ (Ellison–Ellison [11, p.258, l.11]).

(5) Suppose a desired relation can only be arrived at tentatively. That is, we have an effective construction of the relation only for special cases and it is difficult to find an *effective* construction for the general case. However, after assuming the validity of the relation, we can develop an interesting theory for the general case. Consequently, it is necessary to assume the relation's existence if we want to discuss the development.

Example 2.14.

Suppose we want to see if the argument in the proof of Edwards [10, p.147, Chinese Remainder Theorem] can be applied to a commutative ring (Ireland–Rosen [21, p.181, Proposition 12.3.1]). Thus, in Ireland–Rosen [21, p.181, Proposition 12.3.1], we assume $A_i + A_j = R$. That is, the existences of u_1 and v_1 in Ireland–Rosen [21, p.181, l.−10] are assumed by hypothesis. In the case of $\mathbb{Z}[e^{2\pi i/\lambda}]$, a long explanation is required to use the fewest possible instances of trial-and-error to construct u_1 and v_1 in a finite number of steps (See Edwards [10, pp.147–149]). However, we cannot simulate the above effective algorithm in a general communicative ring, i.e., in reality u_1 and v_1 for the general case may come from nowhere. The assumption $A_i + A_j = R$ will enable us to use the equality without knowing its origin, keep the context in a reasonable size and focus on the pattern of argument.

Example 2.15.

Edwards [10, p.62, Theorem] ($p = 2 \times 5 + 1$)
$\longrightarrow$ Edwards [10, p.63, Theorem] ($p = 2n + 1$ is prime in many cases)
$\longrightarrow$ Edwards [10, p.64, Sophie Germain's theorem] (The existence of p is assumptive).

Example 2.16. (Ince [20, p.197, l.−15–l.−8])

2.3.1 The existence produced by the axiom of choice

Sometimes we are not sure about the existence of a function, but we use an axiom to assume its existence. Many people believe that in this way we plug the hole of the theory as if the theory originally had a leak. Other people believe that the theory is not completely perfect if we quote a dubious axiom, but in terms of the difficult of understanding, we still remain at the first level of hell, provided we stick to quoting the same axioms even if many times. These are all false concepts to evaluate the quality of a theory. My point is that every time one quotes the same dubious axiom, one leads us to the next level of hell. For example, when one quote such an axiom three times, one actually leads us to the third level of hell, which is worse than the second. If one quotes a dubious axiom indiscriminately in a theory, one leads us perpetually to further descend in hell. One should not construct a function by an axiom if one could do it by an exact formula. The best procedure is to keep the number of times we quote a dubious axiom to a minimum. This is the reason why Weiner's original proof (Wiener [46, p.75, l.−15–p.97, l.7]) of general Tauberian theorem (Wiener [46, p.74, (10.04)] and Rudin [37, p.211, Theorem 9.7(a)]) is clear and valuable but other proofs are not. This is because other proofs use the Hahn–Banach theorem (Rudin [37, p.59, Theorem 3.5]). See Rudin [37, p.212, l.11; p.211, l.13 & l.8]. The proof of Rudin [37, p.59, Theorem 3.5] uses the axiom of choice (Rudin [37, p.59, l.−12; l.13; l.1; p.57, l.11–l.18]).

Dugundji [9] presents the axiom of choice in the first chapter (Dugundji [9], p.21, l.16–l.17]), while Hall–Spencer [16] pesents the same axiom in the last chapter (Hall–Spencer [16, chap. 7]). Because a dubious axiom affects a theory's quality, we should separate the category of theorems that can be derived without using the axiom of choice from the category of theorems that cannot. Dugundji's arrangement makes it difficult for readers to determine to which category a theorem in Dugundji [9] belongs, especially for theorems near the end of the book. In contrast, Hall–Spencer [16] confines all the theorems that involve the axiom of choice to the last chapter. With regard to the reader's convenience, Hall–Spencer [16] has made the better choice.

2.4 Close existence vs. distant existence

If the derivation of an existence requires only a simple procedure, then it is called a close existence; if the derivation of an existence requires a complicated procedure, then it is called a distant existence, whether a constructive existence, an existence derived from reduction to absurdity, or an assumptive existence.

Example 2.17.

The rule for determining the solvability of $Ax = b$ given in Ince [20, p.207, l.26–l.28] requires only simple calculations, while the rule for determining the solvability of $Ax = b$ given in Coddington–Levinson [3, p.294, l.14–l.15] requires a complicated procedure. Compare the following two proofs.

Proof of the sufficient condition for the first rule. rank $(A) =$ rank $(B) \Rightarrow \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix} = \sum_1^M c_i \begin{pmatrix} a_{1i} \\ \vdots \\ a_{Mi} \end{pmatrix}$, where $\begin{pmatrix} a_{1i} \\ \vdots \\ a_{Mi} \end{pmatrix} = (A)e_i$. □

Proof of the sufficient condition for the second rule. Let A be an $m \times n$ matrix, and $A : \mathbb{C}^n \to \mathbb{C}^m$. Then $A^* : \mathbb{C}^m \to \mathbb{C}^n$.

$A^*u = 0 \Leftrightarrow \begin{pmatrix} \overline{a_{11}} & \cdots & \overline{a_{m1}} \\ \vdots & \ddots & \vdots \\ \overline{a_{1n}} & \cdots & \overline{a_{mn}} \end{pmatrix} \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} = 0$

$\Leftrightarrow \sum_{i=1}^m u_i \overline{a_{ij}} = 0 \quad (j = 1, \cdots, n)$

$$\Leftrightarrow (u, \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}) = 0 \quad (j = 1, \cdots, n).$$

Thus, (the null space of A^*) $= (A\mathbb{C}^n)^\perp$, where $\perp$ refers to the scalar product on $\mathbb{C}^m$.
By hypothesis, $b \in$ (the null space of $A^*)^\perp = ((A\mathbb{C}^n)^\perp)^\perp$
$= A\mathbb{C}^n$ [Jacobson [22, vol. 2, p.151, l.17]]. Consequently,
$\exists x \in \mathbb{C}^n : b = Ax$. □

3 How generalization affects existence and uniqueness

A more generalized theorem reveals fewer features. For example, each of the following four statements discusses the existence and uniqueness of factorization.

(1) Van der Waerden [42, vol.1, p.61, l.18–l.20]: an Euclidean domain.

(2) Jacobson [22, vol.1, p.122, Theorem 2]: a principal ideal domain.

(3) Stewart–Tall [41, p.186, Theorem 10.1]: the ring of integers is generated by a single element.

(4) Stewart–Tall [41, p.117, Theorem 5.5]: a Dedekind ring (Stewart–Tall [41, p.116, l.17]).

For 1 and 2, let us restrict our consideration to the case where the ring of integers is generated by a single element. If all the ideals are principal, then 3 reduces to 2. Therefore, each statement is a generalization of the previous statement.

3.1 How generalization affects uniqueness

Let us consider the uniqueness of prime factorization. From 1 to 4, the proof of uniqueness of each statement can be applied to the proof of the previous statement. In other words, the uniqueness in each statement implies the uniqueness of the previous statement even though the uniqueness of each statement can have a proof which relies on its own features rather than the features of a more generalized statement. Here are some examples: van der Waerden [42, vol.1, p.60, l.−3–p.61, l.17] proves the uniqueness in 2; Ribenboim [34, p.268, l.18] proves the uniqueness in 3.

3.2 Counterexamples

Example that satisfies the conditions in 4 but not those in 3: Stewart–Tall [41, §2.6, Example 2(b)].
Examples that satisfy the conditions in 3 but not those in 2: Stewart–Tall [41, p.111, l.1–p.112, l.11; p.113, l.12–l.−1].

These examples provide some valuable ideas about prime factorization (Stewart–Tall [41, p.112, l.12–l.23; p.113, l.1–l.8]).

Remark. In terms of unique prime factorization, statement 2 is a dividing line between elements and ideals (Stewart–Tall [41, p.110, l.−1]): moving from 1 to 2, we find that that a principal ideal domain is the *weakest*

domain that allows unique factorization (i.e., a unique factorization domain (Stewart–Tall [41, p.96, l.17]) must be a principal ideal domain) (Stewart–Tall [41, p.132, Theorem 5.15]); moving from 4 to 2, we find that the unique factorization is universal (i.e. valid for ideals) and that statement 2 just happens to be the *special* case when all the ideals are principal (i.e. ideals become elements).
Examples that satisfy the conditions in 2 but not those in 1: $\mathbb{Z}(\sqrt{d})$, where $d = -19, -43, -67, -163$ (Stewart–Tall [41, p.93, l.10; p.101, Theorem 4.18; p.132, Theorem 5.15]).

3.3 How generalization affects existence

From 4 to 1, the construction of existence is more specific and effective because there are more tools and resources available. In fact, statement (i) reveals an extra feature that statement $(i+1)$ cannot $(i = 1, 2, 3)$. For example, 3 recognizes that the prime divisors of p are conjugate to each other. The proof of Stewart–Tall [41, p.117, Theorem 5.5] uses the axiom of choice in Stewart–Tall [41, p.117, l.-4] (See Dugundji [9, p.31, Theorem 2.1] and the proof of McCoy [30, p. 454, Theorem 17.32]), while Stewart–Tall [41, p.186, Theorem 10.1] is proved without using the ineffective axiom of choice.
Remark. In my opinion, if we can find a more effective method than using the axiom of choice to prove the unique factorization for some special cases that satisfy the conditions in Stewart–Tall [41, p.117, Theorem 5.5], then we should preserve these proofs because effective methods are irreplaceable by an ineffective method.

4 Uniqueness

4.1 How we refine the proof of the uniqueness theorem

We may refine a theorem by weakening its hypothesis. 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. 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. 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. For details, read Wang [44, §4].

In the following ordered set of uniqueness theorems about generalized Lipschitz conditions, each theorem is more refined than its previous theorems:

(1) Coddington–Levinson [3, p.10, Theorem 2.2]

(2) Coddington–Levinson [3, pp.48–49, Theorem 2.1]

(3) Coddington–Levinson [3, p.49, Theorem 2.2]

(4) Coddington–Levinson [3, p.51, Theorem 2.3]

Remark 1. Coddington–Levinson [3, p.49, l.13–l.20; p.51, l.-12–l.-5]

Remark 2. The more refined a theorem is, the less effective its proof is. It may require frequent use of reduction to absurdity to prove a refined theorem.

Remark 3. Both Jackson [23, p.37, l.−13–p.38, l.7] and Conway [6, p.255, Corollary 1.9] prove the uniqueness of the solution to the Dirichlet problem [Sneddon [39, p.151, l.−9]]. The former proof uses Green's first identity [Jackson [23, p.36, (1.34)]], while the latter proof uses the mean value property [Conway [6, p.253, Definition 1.5]]. However, the former proof can be used to prove the uniqueness of the solution to the Neumann problem at the same time.

4.2 How we find more effective methods to prove the same uniqueness theorem

In order to effectively prove a theorem, we should fully utilize the resources. If we try to quote a theorem and a special case is sufficient, we should not quote the general case. My reason are as follows: First, this quotation is customized to our needs and reduces the gap between the quoted theorem and the theorem to be proved by establishing more close relationships between them. Second, the method helps focus on our goal. Third, the method enables us to become familiar with the surroundings of the theorem to be proved; the ambient features may inspire our proof strategy. Quoting a theorem is like purchasing clothes: quoting the general case is like choosing the thickness of clothes; quoting a special case is like choosing their size and thickness. If we know more about a theorem's ambient features, we know more about the theorem. Fourth, the precious part of problem solving is its process instead of its final answer.

Example 4.1. (The uniqueness theorem for the initial value problem $x' = f(t,x)$, where f satisfies the Lipschitz condition)

The order of the following list of proofs is based on the effectiveness of their arguments:

(1) Coddington–Levinson [3, p.10, Theorem 2.2] follows from Coddington–Levinson [3, p.8, (2.2); case $\varepsilon = 0$].

(2) Birkhoff–Rota [2, p.142, Theorem 1] follows from Birkhoff–Rota [2, p.24, Lemma 2].
Note that Birkhoff–Rota [2, p.24, Lemma 2] is a generalization of Coddington–Levinson [3, p.8, (2.2); case $\varepsilon = 0$].

(3) Collatz [5, chap. I, §6, Sec.19]
Advantage: It can be generalized to n-dimensional vector spaces.
Shortcoming: It uses the method of reduction to absurdity.

(4) Ince [20, p.65, l.17–p.66, l.7]; Hartman [18, p.9, l.−5–p.10, l.2]
Shortcoming: Both quoted passages unnecessarily use the method of mathematical induction.

(5) Pontryagin [35, p.157, l.11–p.158, l.−7]
Criticism: If we compare it with 1 or 2, it contains no new ideas except adding the unnecessary topological mechanism to the argument. Topology is useful only in illuminating mathematical structures.

4.3 Sufficient conditions for uniqueness

The *coefficients* of the power series of the solution for Cauchy's problem are *uniquely* determined by Petrovsky [33, p.18, (9.2) & (14,2)]. See Petrovsky [33, p.18, l.−2–p.20, l.20]. Therefore, the solution for Cauchy's problem is unique. In Dirichlet's problem (John [24, p.95, l.9–l.10]), the boundary conditions imply uniqueness (John [24, p.95, l.7]). For the heat equation (John [24, p.217, (1.36a)]), the initial and growth conditions imply uniqueness. John [24, pp.215–217, three theorems] show these initial and growth conditions *can* be developed from the boundary conditions. In John [24, p.139, l.−3–p.140, l.−14], the *energy* integral is used to prove the uniqueness of a mixed problem for the wave equation. This observation provides strong support for the use of a system's *total energy* in establishing the equation of motion as a postulate in quantum mechanics (Cohen-Tannoudji–Diu–Laloë [4, vol. 1, p.222, Sixth postulate]).

4.4 Why the uniqueness of the solution of Cauchy's problem can fail

Example 4.2.

(1) The hypothesis fails to satisfy a Lipschitz condition (Ince [20, p.66, l.−10–l.−8; p.67, l.1–l.21]; Birkhoff–Rota [2, p.23, l.1–l.2]).

(2) The hypothesis fails to satisfy a growth condition (John [24, p.217, l.12–l.15]).

(3) The Cauchy data (John [24, p.245]) is outside the cone on which the initial conditions determine a unique solution (Petrovsky [33, p.76, l.−9–p.77, l.8]). The example given in Petrovsky [33, p.76, l.−9–p.77, l.8] provides a systematic method for constructing a family of solutions of the Cauchy problem.

(4) Characteristic surfaces impede the coordinate transformation from the generalized Cauchy problem to the standard Cauchy problem (Petrovsky [33, p.29, l.22–l.24; p.30, l.17–p.31, l.2]).

4.5 How we preserve the convergence of successive approximations when generalizing the Lipschtiz condition

Read Coddington–Levinson [3, p.54, Theorem 3.1].

4.6 How to strengthen the argument for uniqueness

(1) Allowing more solutions to be included.
Analytic functions (Petrovsky [33, p.20, l.18–l.20]) $\longrightarrow$ non-analytic functions (Petrovsky [33, p.34, §4]).

(2) Weakening the condition that ensures uniqueness.
The growth condition (John [24, p.217, (1.36c)]) $\longrightarrow$ the unilateral condition (John [24, p.222, (1.57c)]). See John [24, p.222, l.9–l.14].

(3) Using general properties instead of special ones.
Petrovsky [33, p.75, §11] shows that we may use Green's theorem to prove uniqueness of the solutions of wave equations. By John [24, p.129, (1.14)], it is unnecessary to go that far.

4.7 Generalizations derived from uniqueness

(1) The integration of differential inequalities: Hartman [18, p.24, Exercise 1.1] $\longrightarrow$ Hartman [18, p.24, Theorem 1.1] $\longrightarrow$ Wikipedia [47, §Integral form for continuous functions, (a) and (b)].

(2) The upper bound and lower bound of solutions' absolute values: Hartman [18, p.27, Corollary 4.3].

5 Conclusion

In order to keep the methods in a mathematical theory effective, in most cases we should prioritize our choices in the following order: constructive existence, the existence derived from reduction to absurdity, and assumptive existence.

Based on this guideline, Hardy–Wright [17, p.162, Theorem 192] is better than Hardy–Wright [17, p.160, Theorem 190]. The existence derived from reduction to absurdity is often used in the general case when a general formula for constructive existence is not available. For example, in order to present a theory, we are forced to use existences derived from reduction to absurdity in Davenport [8, p.117, l.9 & l.-10], but in a numerical case (Davenport [8, p.118, l.-12–p.119, l.17]), we replace an existence derived from reduction to absurdity with a constructive existence.

In Davenport [8, p.120, l.17–l.19], Davenport says, "A construction often gives greater mental satisfaction than a mere proof of existence, though the distinction between the two is not always a clear-cut one." For example, the difference between a constructive existence and an existence derived from reduction to absurdity depends on how much indefiniteness the existence derived from reduction to absurdity may cause in that particular case. If the set of indefinite choices has only one element, the two concepts are the same; otherwise, they are different. As another example, the existence produced by the axiom of choice is assumptive, but the existence is also constructive in a weak sense. See 2.3 (1). Even though an existence may fall into the overlap of two categories, this confusing and complicated nature should neither prevent us from studying its origin and intrinsic nature nor mislead us to be indiscriminate about various types of existences.

References

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

[2] Birkhoff, G. & Rota, G. C.: *Ordinary Differential Equations*, 3rd ed., New York: John Wiley and Sons, 1978.

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

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

[5] Collatz, L.: *Differential Equations*, translated from German by E. R. Dawson, New York: John Wiley & Sons, 1986.

[6] Conway, J. B.: *Functions of One Complex Variable*, 2nd ed., New York: Springer-Verlag, 1978.

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

[8] Davenport, H.: *The Higher Arithmetic*, 5th ed., Cambridge: Cambridge University Press, 1982.

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

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

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

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

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

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

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

[16] Hall, D. W. & Spencer II, G. L.: *Elementary Topology*, New York: John Wiley & Sons, 1955.

[17] Hardy, G.H. & Wright, E.H.: *An Introduction to the Theory of Numbers*, 5th ed., Oxford: Clarendon Press, 1979.

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

[19] Hua, L.K.: *Introduction to Number Theory*, Translated from the Chinese by Peter Shiu, New York: Springer-Verlag, 1982.

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

[21] Ireland, K. & Rosen, M.:*A Classical Introduction to Modern Number Theory*, 2nd ed., New York: Springer-verlag, 1990.

[22] Jacobson, N.: *Lectures in Abstract Algebra*, vol.1, Princeton: Van Nostrand, 1951.

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

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

[25] Johnson, R.E., Kiokemeister F.L., and Wolk, E.S.: *Calculus*, New York: 2nd ed., Boston: Allyn and Bacon, Inc., 1971.

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

[27] Karamcheti, K.: *Vector Analysis And Cartesian Tensors With Selected Applications*, San Francisco: Holden-Day, 1967.

[28] Landau, E. G. H.: *Foundations of Analysis*, translated from the German by F. Steinhardt, New York: Chelsea Publishing Company, 1951.

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

[30] McCoy, N. H.: *Fundamentals of Abstract Algebra*, Boston: Allyn and Bacon, 1972.

[31] Niven, I. & Zuckerman, H.S.: *An Introduction to the Theory of Numbers*, 4th ed., New York: John Wiley & Sons, 1980.

[32] Pervin, W.J.: *Foundations of General Topology*, New York, Academic Press, 1964.

[33] Petrovsky, I. G.: *Lectures on Partial Differential Equations*, (translated from Russian by A. Shenitzer), 1st English ed., New York: Interscience Publishers, 1954.

[34] Ribenboim, P: *Algebraic Numbers*, New York: Wiley-Interscience, 1972.

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

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

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

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

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

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

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

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

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

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

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

[46] Wiener, N.: *The Fourier Integral and Certain of Its Applications*, New York: Dover Publications, 1951.

[47] Wikipedia: *Gronwall's inequality*, `http://en.wikipedia.org/wiki/Gronwall'sinequality`

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