The Arsenal of Perturbation Bounds for Finite Continuous-Time Markov Chains: A Perspective

Alexander Y. Mitrophanov

Frederick National Laboratory for Cancer Research, National Institutes of Health, Frederick, MD 21702, USA; alex.mitrophanov@nih.gov or mitrophanov.alex@gmail.com

Abstract: Perturbation bounds are powerful tools for investigating the phenomenon of insensitivity to perturbations, also referred to as stability, for stochastic and deterministic systems. This perspective article presents a focused account of some of the main concepts and results in inequality-based perturbation theory for finite state-space, time-homogeneous, continuous-time Markov chains. The diversity of perturbation bounds and the logical relationships between them highlight the essential stability properties and factors for this class of stochastic processes. We discuss the linear time dependence of general perturbation bounds for Markov chains, as well as time-independent (i.e., time-uniform) perturbation bounds for chains whose stationary distribution is unique. Moreover, we prove some new results characterizing the absolute and relative tightness of time-uniform perturbation bounds. Specifically, we show that, in some of them, an equality is achieved. Furthermore, we analytically compare two types of time-uniform bounds known from the literature. Possibilities for generalizing Markov-chain stability results, as well as connections with stability analysis for other systems and processes, are also discussed.

Keywords: Markov stochastic process; Kolmogorov equation; differential equation; perturbation theory; sensitivity analysis; stability; robustness; ergodicity coefficient; stationary distribution

MSC: 60J27; 60J35; 34D10; 60J28; 34D20

1. Introduction

Perturbation bounds and related approaches for continuous-time Markov chains have been applied in research fields as diverse as reliability theory [1–3], queuing theory [4–8], quantum physics [9–12], climate science [13], biochemical kinetics [14–19], economics [20], population genetics [21], and health insurance modeling [22]. In principle, such bounds can be useful in any field where continuous-time Markov chains and their generalizations are used as mathematical models. At the same time, Markov chain perturbation bounds represent noteworthy theoretical developments that have connections with many directions of mathematical research. In this perspective article, we will summarize and highlight some distinguishing features of Markov chain perturbation bounds that illustrate both the inner logic of this research area and its usefulness for current and future applications. Specifically, we will discuss exponential vs. linear time dependence for perturbation bounds, as well as their possible time-independence (or, time-uniformity) and the connection with the rate of exponential convergence to the stationary distribution. Moreover, we will provide new results characterizing the tightness of time-uniform perturbation bounds. Additionally, we will outline the relationships between different perturbation-theory results for Markov chains and other processes and systems.

Perturbation bounds, their properties, and the connections between them constitute inequality-based perturbation theory, which can be developed for Markov chains and, generally, for stochastic and deterministic processes (i.e., mathematical objects representing systems changing over time). This complements the more traditional approach to perturbations that focuses on continuity and differentiability results, as well as asymptotic
expansions [23–28]. For quantitative studies, both perturbation bounds and perturbation expansions have their respective advantages; a comparison between them has been attempted in the case of discrete time [29]. For the purposes of this article, we emphasize that Markov chain perturbation bounds provide (1) a compact and convenient representation of the essential features defining the chain’s sensitivity to perturbations and (2) a bound for the magnitude of the perturbation in the Markov chain’s state probabilities given the magnitude of the perturbation in the chain’s parameters and initial distribution. This magnitude is often a “summary” of the perturbation magnitudes for the chain’s individual parameters, which allows the bound to hold for perturbations of different structure but the same magnitude. Note that, while the discovery of informative lower perturbation bounds would be very insightful, current research focuses on upper bounds, which is what we discuss in this article.

The primary reason for the focus on continuous time (in this perspective article, as well as in most of the author’s research) is that physical time is continuous. This makes continuous-time Markov chains a natural choice for the stochastic modeling of real-world phenomena and systems. One prominent example is provided by physics and chemistry, where the (forward) Kolmogorov equations, which govern temporal changes in the Markov chain’s state probabilities, have a special name: the master equation [15,30]. Yet another reason is the close connection with a powerful branch of mathematics—the theory of differential equations. Indeed, the Kolmogorov equations are a system of differential equations. One could thus anticipate that the general perturbation theory for differential equations would guide us toward the perturbation bounds we need. One of the well-known results in differential-equation theory is Gronwall’s inequality and its different versions [31,32]. The application of this inequality to Markov chains (which has been attempted more than once, including an article in this Special Issue [22]) is what motivated us to write this perspective article.

Herein, we consider finite, time-homogeneous chains, because they provide excellent opportunities for illustrating the main concepts of perturbation analysis and also due to the considerable importance of such chains for applications. The possibility of generalizations to countable state spaces and time-inhomogeneous Markov chains will be indicated in the comments. Furthermore, this article focuses on regular perturbations, which correspond to cases where expected perturbation magnitudes can be regarded as small. This smallness is, often, not a strict mathematical requirement but a reflection of situations where such bounds can be useful. In contrast, singular perturbations correspond to cases where some state transitions in a Markov chain are considerably faster than others, so we could think of “large-magnitude perturbations” or multiple time scales. While the typical approach to singular perturbations centers on asymptotic expansions [25,26,28], perturbation-bound approaches to singular perturbations have also been developed [33,34]. Thus, some of the results that we discuss could, in principle, be applied to singular-perturbation problems.

This perspective article describes what can be regarded as deterministic perturbations of the Kolmogorov equations. Thus, we are in effect considering deterministic perturbations of a stochastic process (i.e., the Markov chain under study). One could possibly imagine perturbation scenarios involving various deterministic or stochastic systems under deterministic or stochastic perturbations. Clearly, each scenario would require its own theoretical developments. Yet, the types of results we discuss could be relevant in a broader context and may be applicable to other possible (and, as it might happen, far more complex) perturbation scenarios. At the very least, they can provide a relevant standard for comparison or even help generate a viable working hypothesis [35,36].

2. Continuous-Time Markov Chains and Perturbations: Notation and Some Basic Properties

Let \( S = \{0, 1, \ldots, N\} \), where \( N \geq 1 \) is an integer, be a finite set. On this set, regarded as the state space, consider a continuous-time Markov chain \( X = \{X(t), t \geq 0\} \) with constant generator (also known as the transition-rate matrix) \( Q = (q_{ij}) \) and vector of state probabilities \( p(t) = (p_i(t)) \). Here, \( q_{ij} \) is the rate of transitions from state \( i \) to state \( j (i \neq j) \),
and \( p_i(t) \) is the probability that \( X(t) \) will be in state \( i \), given an initial distribution \( p(0) \) (see, e.g., the definition of a continuous-time Markov chain in Refs. [37,38]). On the same state space, consider another Markov chain, \( \tilde{X} = \{ \tilde{X}(t), t \geq 0 \} \), with generator \( \tilde{Q} = \{ \tilde{q}_{ij} \} \) and vector of state probabilities \( \tilde{p}(t) = (\tilde{p}_i(t)) \). We will refer to the chains \( X \) and \( \tilde{X} \) as the unperturbed and perturbed chains, respectively, and the matrix \( E := \tilde{Q} - Q \) is the perturbation. To measure the magnitude of perturbations, we will use the \( l_1 \) norm (absolute entry sum) for vectors, which will be regarded as row vectors (per the tradition existing in the Markov chain literature). For matrices, we will use the corresponding subordinate norm, which is the maximum absolute row sum. We will denote \( l_1 \) vector and matrix norms by \( \| \cdot \| \). Thus, for a vector \( x = (x_i) \) and a matrix \( A = (a_{ij}) \), the norms are defined as follows:

\[
\| x \| = \sum_i |x_i|, \quad \| A \| = \sup_{|x| = 1} \| xA \| = \max_i \sum_j |a_{ij}|.
\]

Importantly, for differences between probability vectors, this choice of norm corresponds to variation distance, which arguably is the most widely used distance in contemporary Markov chain theory (at least in the case of finite state spaces). For probability vectors \( p \) and \( \tilde{p} \) representing distributions on \( S \), the variation distance, \( d_{TV}(., .) \), is commonly defined as follows:

\[
d_{TV}(\tilde{p}, p) := \sup_{A \subseteq S} \| \tilde{p}(A) - p(A) \| = \frac{1}{2} \| \tilde{p} - p \|,
\]

where \( \tilde{p}(A) \) and \( p(A) \) are the measures on \( S \) induced by \( \tilde{p} \) and \( p \), respectively. Sometimes, the quantity \( 2d_{TV}(\tilde{p}, p) = \| \tilde{p} - p \| \) is used as the variation distance [18,19]. Because any two norms in a finite-dimensional space are equivalent, a bound on \( \| \tilde{p} - p \| \) will imply a bound on the vector difference in any other norm of interest, but the corresponding absolute constant may not be readily available. And even if it is, the resulting bound may not be tight (i.e., it might considerably overestimate the actual perturbation magnitude). A preferred approach would be to follow the proof of a bound in the \( l_1 \) norm and see if the same proof, perhaps with small modifications, also works for another norm of interest, such as an \( l_p \) norm (see, e.g., Ref. [39]).

Define \( z(t) := \tilde{p}(t) - p(t) \), so \( \| z(t) \| \) is the magnitude of the perturbation in the state-probability vector of the chain \( X \) at time \( t \geq 0 \). To avoid the trivial case, we will assume throughout the article that \( E \neq 0 \); this assumption is necessary for some of the perturbation bounds to be strict inequalities. However, cases where \( z(0) = 0 \) will not be excluded. The perturbation bounds that we discuss will typically be uniform over a (finite or infinite) time interval and have the form

\[
\sup_{t \in [0,T]} \| z(t) \| \leq \kappa_1(T) \| z(0) \| + \kappa_2(T) \| E \|, \quad 0 < T \leq \infty,
\]

where \( \kappa_1(T) \) and \( \kappa_2(T) \) are the condition numbers (this term was borrowed from numerical linear algebra, where perturbation bounds are prevalent [40]). If these numbers are sufficiently small, then the chain \( X \) is well conditioned and insensitive to perturbations. While large condition numbers do not necessarily mean that the chain is sensitive, it is often implied and might as well be true. In any event, for our sensitivity assessment to be accurate, we want the bound in Equation (1) to be as tight as possible.

The use of the \( l_1 \) norm in Equation (1) offers some analytic advantages. First, because \( \| p(t) \| = 1 \) (due to \( p(t) \) being a probability vector), Equation (1) naturally provides a bound involving relative perturbations:

\[
\sup_{t \in [0,T]} \frac{\| z(t) \|}{\| p(t) \|} \leq \kappa_1(T) \frac{\| z(0) \|}{\| p(0) \|} + \kappa_2(T) \| Q \| \frac{\| E \|}{\| Q \|}.
\]

(2)
Equation (2) shows that the perturbation in $p(t)$ will be small if the relative perturbations in $p(0)$ and $Q$ are both sufficiently small. In fact, using the $l_1$ norm, the absolute and relative perturbations of $p(t)$ are equivalent due to $\|p(t)\| ≡ 1$. At the same time, for the chain $X$ to be well conditioned with respect to absolute perturbations in the generator, $\kappa_2(T)$ needs to be sufficiently small, whereas for it to be well conditioned with respect to relative perturbations in the generator, $\kappa_2(T)\|Q\|$ needs to be sufficiently small, as follows from Equation (2).

A second advantage of the $l_1$ norm is that Equation (1) can be divided by $N + 1$ (i.e., the size of the state space) and thereby provide a bound on the average perturbation in a state probability of the chain $X$ (averaged over all state-probability perturbations). For some applications, the metric $\|z(t)\|/(N + 1)$ might be more informative than $\|z(t)\|$ (cf. Ref. [14]). The division by $N + 1$ also allows one to control the growth of the right-hand side of Equation (1) with $N$, which can occur due to the nature of norm-based bounds.

Moreover, the use of the $l_1$ norm allows us to obtain simple perturbation bounds for the moments of the random variable $X(t)$, as demonstrated by the following statement (in which $E(.)$ and $var(.)$ denote expectation and variance, respectively).

**Statement 1.** The following bounds hold for all $t \geq 0$ and every positive integer $m$:

$$\begin{align*}
\left| E\left(\tilde{X}^m(t)\right) - E(X^m(t)) \right| &\leq N^m\|z(t)\|, \\
\left| var\left(\tilde{X}(t)\right) - var(X(t)) \right| &\leq 3N^2\|z(t)\|.
\end{align*}$$

**Proof.** The perturbation bound for the non-central moments is a direct generalization of the corresponding result for the expectation [4]:

$$\left| E\left(\tilde{X}^m(t)\right) - E(X^m(t)) \right| \leq \sum_{k=0}^{N} k^m |\tilde{p}_k(t) - p_k(t)| \leq N^m\|z(t)\|.$$  

Next, from the basic properties of variance, we have

$$\left| var\left(\tilde{X}(t)\right) - var(X(t)) \right| \leq \sum_{k=0}^{N} k^2 \left( \tilde{p}_k(t) - p_k(t) \right) + \left( \sum_{k=0}^{N} k\tilde{p}_k(t) \right)^2 - \left( \sum_{k=0}^{N} kp_k(t) \right)^2.$$  

Here, the first term on the right-hand side does not exceed $N^2\|z(t)\|$. For the second term, we have

$$\left| \sum_{k=0}^{N} k\tilde{p}_k(t) - \sum_{k=0}^{N} kp_k(t) \right|^2 \leq N\|z(t)\| \times 2N.$$  

Putting everything together, we obtain

$$\left| var\left(\tilde{X}(t)\right) - var(X(t)) \right| \leq N^2\|z(t)\| + 2N^2\|z(t)\| = 3N^2\|z(t)\|.$$  

This completes the proof. □

Notice that, in Statement 1, the $m$th non-central moment for both $X$ and $\tilde{X}$ is bounded by $N^m$, and this bound is attained for distributions concentrated in the state $N$. And if, instead of the absolute moment difference, we consider the relative difference $\left| E\left(\tilde{X}(t)\right) - E(X^m(t)) \right| / N^m$, ...
then, from Statement 1, we obtain a perturbation bound without the explicit dependence on $N^m$ on the right-hand side. Similar relative differences can also be considered for the variances, each of which is bounded by $2N^2$.

3. Time Dependence in Perturbation Bounds: From Exponential to Linear

The forward Kolmogorov equations for chains $X$ and $\tilde{X}$ have the following form:

$$\frac{dp(t)}{dt} = p(t)Q, \quad \frac{d\tilde{p}(t)}{dt} = \tilde{p}(t)\tilde{Q}, \quad t \geq 0.$$  

When Gronwall’s inequality is applied to these equations, Equation (1) holds on finite time intervals and takes the following explicit form [1,14,22]:

$$\sup_{t \in [0,T)} \|z(t)\| \leq e^{\|Q\|T} \|z(0)\| + \left(e^{\|Q\|T} - 1\right) \frac{\|E\|}{\|Q\|}, \quad 0 < T < \infty. \quad (3)$$

The right-hand side of Equation (3) tends to $\|z(0)\|$ as $T \to 0$, suggesting that the bound may be informative on short- or moderate-length time intervals. However, for increasing $T$, the right-hand side grows exponentially, which can make the bound arbitrarily loose.

The possibility of obtaining perturbation bounds with a sub-exponential dependence on $T$ was realized quite early [1]. This sub-exponential dependence turns out to be linear. Indeed, the following bound holds [1,14,33]:

$$\sup_{t \in [0,T)} \|z(t)\| \leq \|z(0)\| + T\|E\|, \quad 0 < T < \infty. \quad (4)$$

The derivation of Equation (4) is rather straightforward and relies on the integral representation of $z(t)$ (using the fact that the Kolmogorov equations are linear) together with some standard norm-based bounds. Using simple calculus, one can show that the right side of Equation (4) is smaller than that of Equation (3) for any $\|Q\|$ and $T$ [14]. Overall, replacing the exponential dependence on $T$ with a linear dependence offers tremendous improvements in bound tightness. However, there is another important conceptual difference between Equations (3) and (4). Specifically, in Equation (3), the condition numbers $\kappa_1(T)$ and $\kappa_2(T)$ depend on the parameters of $X$ via $\|Q\|$; in other words, Equation (3) distinguishes between more well-conditioned and less well-conditioned Markov chains. However, Equation (4) does not make that distinction, and its condition numbers are the same for all Markov chains. Ideally, we would like to combine the tightness of the bound in Equation (4) with the chain-specific nature of the bound in Equation (3). How can this be achieved?

One simple and natural strategy involves reflecting chain-specific information in the choice of $T$, which has been suggested in the context of Markov-chain modeling of the frequently encountered biochemical reaction $A + B \rightleftharpoons AB$ (binary-complex formation and dissociation) [14]. That work investigated the nearness between the quadratic (full) and the linear (approximate) model for the reaction, and the latter was regarded as the unperturbed Markov chain. The author used the fact that the expectation of the unperturbed chain, $E(X(t))$, approached its unique stationary state $E(X(\infty))$ exponentially fast, with exponential rate $\mu$ independent of the initial conditions:

$$\Delta(t) = \Delta(0)e^{-\mu t},$$

where $\Delta(t) := E(X(t)) - E(X(\infty))$. Thus, we can define

$$T_\Delta := \frac{1}{\mu},$$
with $T_\Delta$ being the relaxation time for $\Delta(t)$. It is analogous to relaxation times studied in physics (cf. Ref. [41]), and such terminology has also been adopted in Markov chain convergence research.

Quite intuitively, the relaxation time $T_\Delta$ represents a relevant time scale for temporal changes in $X$, so we can set $T = T_\Delta$ in Equation (4) and thereby obtain a chain-specific—rather than generic—perturbation bound with increased tightness. Note, however, that in Markov chain research, relaxation time is typically defined as the inverse of the spectral gap, which is the spectral characteristic of the generator that defines the rate of convergence of a Markov chain to the stationary distribution [37,42]. In the case of continuous time, the spectral gap can be defined as the minimum absolute real part among all the generator’s nonzero eigenvalues [18,19,37]. Notably, when the unperturbed chain is a Prendiville process, which was the case in the binary-complex formation modeling study [14], the spectral-gap definition of the relaxation time coincides with $T_\Delta$ [18]. Whereas the introduction of $T_\Delta$ assumed uniqueness of the steady state, this approach can be extended to situations where the stationary distribution of the unperturbed chain $X$ is not necessarily unique. Indeed, Equation (4), being general, applies to such cases. All we need is a way to assess the range of relevant time scales for the unperturbed chain. This can be achieved, for example, based on subject-matter expertise in the research field where the Markov chains in question are used as mathematical models.

Setting $T = T_\Delta$ in Equation (4) provides a chain-specific value for $\kappa_2(T)$. At the same time, the value $\kappa_1(T) \equiv 1$ in Equation (4) is still generic. As we will see in the next section, in perturbation bounds suitable for very long time intervals, we also have $\kappa_1(T) \equiv 1$. This essentially is a consequence of the requirement that the bound in Equation (1) be uniform over a certain time interval. When this requirement is absent, the equivalent of $\kappa_1(T)$ can tend to 0 in the infinite-time limit. See, e.g., the bound derivation details in Refs. [18,33].

An important question in the development of Markov chain perturbation theory is the generalizability of the results to time-inhomogeneous and infinite state-space chains. The definition of time-inhomogeneity simply involves Markov chain generators that depend on the time variable: $Q(t)$ and $\tilde{Q}(t)$, $t \geq 0$ [4,22,38]. The perturbation bound in Equation (3) was very recently extended to the case of finite, time-inhomogeneous Markov chains [22].

The bound in Equation (4) can be generalized to time-inhomogeneous chains with a countable state space [7]. While the main focus of Ref. [7] is on Markov chains demonstrating various types of infinite-time convergence (termed ergodicity), the finite-time bound, such as Equation (4), holds in the general case. The main necessary condition is that the theory of differential equations in the Banach (specifically, $l_1$) space is applicable, and a requirement for that is that the generators of the chains under consideration should be bounded. It is worth noting, however, that infinite-time convergence results can be extended to chains with unbounded generators, which serve as mathematical models, e.g., in biology [43]. Likewise, developing a perturbation theory for the case of unbounded generators could benefit some applications.

4. From Linear Time Dependence to Time Independence for Ergodic Markov Chains by Using Convergence Bounds

It turns out that bounds of the form as in Equation (4) can sometimes be considerably strengthened. For this, we need to make an additional assumption: throughout the remainder of this article, we will assume that the stationary distribution of $X$ (i.e., a distribution $\pi = (\pi_i)$ satisfying $\pi Q = 0$) is unique. This assumption is not restrictive, because finite, time-homogeneous, continuous-time Markov chains used in applications very often possess this property. For example, in physics and chemistry, this unique stationary distribution can represent the often-studied state of thermodynamic or chemical equilibrium. For stationary-distribution uniqueness, a frequently used sufficient condition is the irreducibility of $X$ (or, in some applications, the positivity of all transition rates $q_{ij}$ ($i \neq j$), which is sufficient for irreducibility), but it is not required for our purposes. What is required is a rigorous (and, preferably, tight) convergence bound for $X$. 

If $X$ has a unique stationary distribution, $\pi$, then there exist positive numbers $C, b$ such that, for all initial distributions $p(0)$ and $t \geq 0$, we have

$$\|p(t) - \pi\| \leq Ce^{-bt}. \quad (5)$$

This convergence to a unique stationary distribution is the manifestation of ergodicity. Importantly, Equation (5) implies that $C \geq 1$ [19]. If Equation (5) holds, then for all initial distribution vectors $p_1(0)$ and $p_2(0)$ and all $t \geq 0$, we have

$$\|p_1(t) - p_2(t)\| \leq 2Ce^{-bt}, \quad (6)$$

where $p_1(t)$ and $p_2(t)$ are the distributions of $X(t)$ corresponding to the initial distributions $p_1(0)$ and $p_2(0)$, respectively. In the finite, time-homogeneous case, Equations (5) and (6) are equivalent convergence conditions, and they can be proven, e.g., using the properties of the $l_1$ ergodicity coefficient (also known as Dobrushin’s ergodicity coefficient) for chain $X$ [33]. However, Equation (6) is particularly convenient for generalizing perturbation and convergence results to the time-inhomogeneous case. That is why some perturbation results in the literature explicitly use a convergence bound in the form given by Equation (6). The $l_1$ ergodicity coefficient, $\tau_1(\cdot, \cdot)$, is defined for any real square matrix $A$ as follows:

$$\tau_1(A) := \sup_{\|x\| = 1} \|xA\| = \frac{1}{2} \max_{ij} \left| \sum_k a_{ik} - a_{jk} \right|,$$

where $e = (1 \ 1 \ldots \ 1)$ and $^T$ denotes transpose. For a continuous-time chain $X$, ergodicity coefficients are applied to, and calculated for, the chain’s transition matrices $P(t) := \exp(Qt)$.

Since the 1990s, explicit and computable Markov chain convergence bounds have been an active research topic, and numerous such bounds have been obtained for the finite state-space case [16,18,37,44–46]. Their utility in perturbation analysis follows from Equation (5). Indeed, for any $x > 0$, define the mixing time, $\theta(x)$, as follows [37]:

$$\theta(x) := \inf_{t \geq 0} \left\{ t : \|p(t) - \pi\| \leq x \quad \text{for all } p(0) \right\}.$$

For any extent of convergence to the stationary distribution (i.e., for any distance from the stationary distribution), the mixing time $\theta(x)$ is the time when this extent of convergence is achieved. To define a characteristic time of convergence (which is meaningful yet arbitrary), let us choose $x = 1/e$. From Equation (5), it follows that

$$\theta\left(\frac{1}{e}\right) \leq b^{-1}(\log{C} + 1). \quad (7)$$

Combining this with Equation (4), we obtain

$$\sup_{t \in [0,T]} \|z(t)\| \leq \|z(0)\| + b^{-1}(\log{C} + 1)\|E\|, \quad T = b^{-1}(\log{C} + 1). \quad (8)$$

We could have obtained a simpler version of this bound if we had used the relaxation time definition $T = 1/b$ in Equation (4). However, the relaxation time is only a proxy for the mixing time [42], whereas Equation (7) provides a rigorous bound for it. The main reason to use the right-hand side of Equation (8), however, is not the rigor of the mixing-time estimate. Rather miraculously, it turns out that the right-hand side of Equation (8) provides a perturbation bound that is uniform over $t \geq 0$:

$$\sup_{t \geq 0} \|z(t)\| \leq \|z(0)\| + b^{-1}(\log{C} + 1)\|E\|; \quad (9)$$
moreover, this inequality is strict for $C > 1$ [18,19]. Two sufficient conditions for $C > 1$ are that: (1) $N > 1$ and (2) $N = 1$ and the stationary distribution of the chain $X$ is non-uniform [19]. Thus, $C > 1$ is—by far—the prevalent case.

Besides the obvious significance of the time-uniform bound, such as Equation (9), in the analysis of regular perturbations, this time-uniformity is essential in the derivation of bounds for singular perturbations [33]. Equation (9) eliminates the time-dependency on the right-hand side altogether, while preserving the chain-specific nature of the bound. This bound delivers a clear message: if, for a Markov chain, we have a convergence bound of the type shown in Equation (5) (or Equation (6)), then we “automatically” obtain a perturbation bound for that Markov chain. Moreover, (1) if a chain converges fast to its stationary distribution, then it is stable under perturbations in its generator, and (2) the main determinant of this stability is the exponential convergence rate $b$. Thus, obtaining perturbation bounds is another reason why mathematicians should study Markov chain convergence, which complements the list of such reasons given in the preface to the first edition of Ref. [37].

Because the derivation of Equation (9) yields a bound that holds on an infinite time interval, that approach also works for stationary distributions. Indeed, if $\tilde{\pi}$ is a (not necessarily unique) stationary distribution of $\tilde{X}$, then

$$\|\tilde{\pi} - \pi\| \leq b^{-1}(\log C + 1)\|E\|;$$

(10)

if $C > 1$, then this inequality is strict [18]. The perturbation bounds in Equations (9) and (10) were derived specifically for the Kolmogorov equations and use some unique features of their solutions. They rely on the notion of the ergodicity coefficient, which plays an important role in the theory of stochastic matrices [39,40,47]. (See Ref. [33] for a perturbation analysis in continuous time with more extensive use of ergodicity coefficients.) These bounds illustrate how, by exploiting the special structure of the governing equations for different classes of stochastic (and deterministic) processes, one can obtain increasingly informative and accurate perturbation and approximation results.

The strictness of the inequalities in Equations (9) and (10) for $C > 1$ helps us to avoid the futile, in this case, search for examples of equality, which the non-strict inequality in Equation (8) could encourage (and which, in general, can be very meaningful for a perturbation bound). At the same time, if the bounds in Equations (9) and (10) turned out to be strict for all possible $C$, including $C = 1$, then we would have been motivated to try to improve these bounds using an absolute multiplicative constant (which is another meaningful pursuit in general perturbation theory). However, this is impossible, as demonstrated by the following statement.

**Statement 2.** There exist two-state Markov chains $X$ and $\tilde{X}$ for which, in Equations (9) and (10), an equality is attained.

**Proof.** First, consider Equation (10), which is non-strict for $N = 1$, suggesting that, in this case, an equality is possible. Choose $N = 1$ and choose $Q$ so that $q_{01} = q_{10} = 1$ (the other two entries of $Q$ are determined from the condition that row sums for any generator $Q$ are all equal to 0). Due to this symmetry, the stationary distribution of $X$ is uniform, i.e., $\pi = (1/2 \ 1/2)$. Now, on the same state space $S = \{0, 1\}$, choose $\tilde{Q}$ so that $\tilde{q}_{01} = 1$ and $\tilde{q}_{10} = 0$. Obviously, the corresponding stationary distribution is unique and equal to $\tilde{\pi} = (0 \ 1)$. Direct calculation shows that $\|E\| = 2$ and $\|\tilde{\pi} - \pi\| = 1$. The chain $X$ is a special case of the Prendiville process on $S = \{0, 1\}$, and an explicit formula for its $l_1$ ergodicity coefficient, $\beta(t)$, is known [18]:

$$\beta(t) = e^{-(q_{01} + q_{10})t} = e^{-2t}.$$
It can be demonstrated (see Refs. [18,33]) that
\[ \|p_1(t) - p_2(t)\| \leq \beta(t)\|p_1(0) - p_2(0)\|, \]
where \(p_1(t)\) and \(p_2(t)\) are the distributions of \(X(t)\) corresponding to arbitrary initial distributions \(p_1(0)\) and \(p_2(0)\), respectively. Using this inequality together with the general bound \(\|p_1(0) - p_2(0)\| \leq 2\), we obtain
\[ \|p_1(t) - p_2(t)\| \leq 2e^{-2t}. \]

Therefore, we can choose \(C = 1\) and \(b = 1/2\) in Equation (6). We thus have an explicit expression for every quantity on both sides of Equation (10). Substituting them all into that non-strict inequality, we obtain an equality.

Now, for the chosen chains \(X\) and \(\tilde{X}\), and the chosen \(C\) and \(b\), suppose that Equation (9) is a strict inequality. Additionally, assume that \(\|z(0)\| = 0\). Then, from Equations (9) and (10), we have
\[ \sup_{t \geq 0} \|z(t)\| < b^{-1}(\log C + 1)\|E\| = 1 = \|\pi - \pi\| = \lim_{t \to \infty} \|z(t)\|, \]
which is a contradiction. \(\Box\)

5. Related Results and Extensions

Equation (9) is not the only time-uniform bound with a logarithmic dependence on \(C\) reported in the literature. Even though it was the first to be published [18], another bound had, in fact, been derived (and submitted for publication) earlier [19]:
\[ \sup_{t \geq 0} \|z(t)\| \leq \|z(0)\| + \inf_{0 < y < 1} \frac{b^{-1}\log(C/y)}{1-y}\|E\|. \quad (11) \]

The question then becomes, which bound is sharper—Equation (9) or Equation (11)? The following statement shows that the bound provided by Equation (9) is sharper than the one provided by Equation (11).

**Statement 3.** If \(C > 1\), then
\[ \log C + 1 < \inf_{0 < y < 1} \frac{\log(C/y)}{1-y}, \]
and this expression becomes an equality for \(C = 1\).

**Proof.** First, assume that \(C > 1\). Define
\[ f_C(y) := \frac{\log(C/y)}{1-y}. \]

It follows that the infimum of \(f_C(y)\) over \(y \in (0, 1)\) is attained at an internal point \(y_0\) of this interval, such that
\[ C = y_0 \exp\left( \frac{1 - y_0}{y_0} \right) \quad (12) \]
(Proposition 1 in Ref. [19]). Taking the logarithm of Equation (12), we obtain
\[ \log C + 1 = \log y_0 + \frac{1}{y_0}. \]
Therefore, because $0 < y_0 < 1$, we have that $\log y_0 < 0$ and thus
\[
\log C + 1 < \frac{1}{y_0}.
\]
At the same time, from the definition of $f_C(y)$ and Equation (12), it follows that
\[
f_C(y_0) = \frac{1}{y_0}.
\]
This equality, together with the inequality preceding it, proves Statement 3 for $C > 1$. If $C = 1$, then $f_C(y)$ monotonically decreases on $(0, 1)$ and approaches 1 as $y \uparrow 1$ (Proposition 1 in Ref. [19]), from which Statement 3 follows. 

Equation (9) and the related bounds reflect the fact that fast convergence to the stationary distribution implies insensitivity to perturbations. Intuitively, the chain $X$ will be fast converging if all the transition rates in $Q$ are sufficiently large. This begs the question: is it possible to obtain a perturbation bound with the condition number expressed explicitly in terms of the transition rates, $q_{ij}$? This question has been answered in the affirmative for cases where certain additional assumptions are satisfied [19]. A particularly simple answer exists for Markov chains possessing a strongly accessible state, i.e., a state that can be reached from every other state in one transition. An example of such a chain is one whose transition rates $q_{ij}$ ($i \neq j$) are all positive. If $X$ has a strongly accessible state, then, in Equation (1), we can set $T = \infty$, $\kappa_1(T) = 1$, and $\kappa_2(T) = 1/\delta$, where $\delta$ is the sum, over all columns, of the off-diagonal column-minimum entries of $Q$ [19].

A related question is: if the exponential-convergence parameter, $b$, in Equations (9) and (10) is so influential, then could it be possible to obtain perturbation bounds where the condition number $\kappa_2(\infty)$ depends only on $b$ or a related quantity? It turns out that the quantity $N/\lambda$, where $\lambda$ is the spectral gap of $Q$, is a condition number for perturbations of the stationary distribution [19]. (Generally, we have that $b \leq \lambda$ [19], but this can become an equality in many practical situations). A time-uniform perturbation bound with a condition number of $\delta e N/((e - 1)\lambda)$ has been obtained under the additional assumption that the unperturbed chain $X$ is reversible (i.e., is an irreducible chain such that $\pi_i q_{ij} = \pi_j q_{ji}$ for all $i, j \in S$) [48]. Reversible chains form a special class of time-homogeneous Markov chains that is important in many applications, such as models of physical processes that possess the property of detailed balance [18,37,42,48,49].

Equation (9) appears to possess all of the desired perturbation-bound attributes. However, upon a closer look, we may find that there is room for improvement. Indeed, whereas $\|z(t)\| \to \|z(0)\|$ for $t \to 0$, the right-hand side of Equation (9) does not approach $\|z(0)\|$ for small $t$, because it does not depend on $t$ at all. Therefore, the bound in Equation (4), and even Equation (3), can be sharper than Equation (9) for small $t$. How can we handle this situation? Evidently, the only way to maximize tightness is to use Equation (9) on long time intervals and Equation (4) on short ones. For an ergodic unperturbed chain $X$, the inequality in Equation (4) is strict [18,33]. Thus, from Equations (4) and (9), for the prevalent case $C > 1$ we have

\[
\sup_{t \in [0,T]} \|z(t)\| < \begin{cases} 
\|z(0)\| + T \|E\| & \text{if } 0 < T \leq b^{-1}(\log C + 1), \\
\|z(0)\| + b^{-1}(\log C + 1) \|E\| & \text{if } b^{-1}(\log C + 1) \leq T \leq \infty.
\end{cases}
\]
An alternative way to improve Equation (9) via a combination of bounds allows one to handle cases differing in the balance between the magnitude of \( z(0) \) and that of \( E \). Specifically, for \( C > 1 \), the following bound holds [18]:

\[
\sup_{t \geq 0} \| z(t) \| \leq \begin{cases} 
  b^{-1}(\log C + 1)\| E \| & \text{if } \| E \| \geq b\| z(0) \|, \\
  \| z(0) \| + b^{-1}(\log C)\| E \| & \text{otherwise.}
\end{cases}
\]

The intriguing property of the bound above is that, for \( b\| z(0) \| \) small enough relative to \( \| E \| \), the dependence of the right-hand side on \( \| z(0) \| \) disappears altogether.

Analogous bounds to Equation (9), including the limiting distributions (as \( t \to \infty \)), have been derived for time-inhomogeneous Markov chains on finite and infinite state spaces [4–6,50], and convergence bounds for such general cases are also available [16,51]. It turns out, however, that obtaining explicit convergence bounds, such as Equation (5), in the case of an infinite state space (i.e., requiring uniform ergodicity in continuous time; cf. Ref. [52]) can be problematic. Actually, the unperturbed Markov chain of interest may not even be uniformly ergodic. An alternative strategy is to use perturbation bounds that rely on specially selected classes of norms other than the \( l_1 \) norm (such as weighted norms), and this is an active research direction in Markov chain theory and applications [4,5,8,53–56].

6. Discussion

This perspective article is about the properties of perturbation bounds for state-probability vectors of continuous-time Markov chains. All in all, we find these properties rather remarkable. Yet, perhaps the most remarkable of them is the availability and richness of connections with perturbation theories for other classes of quantities, processes, and systems. For example, a certain choice of the pre-exponential factor in a convergence bound—i.e., the constant \( C \) in Equations (9) and (10)—can also be a condition number for the eigenvalues and, therefore, for the spectral gap of the generator \( Q \) [13,18]. One and the same quantity, expressed in terms of the ergodicity coefficients of chain \( X \), can be used as a condition number for the chain’s state-probability vectors and also for its ergodicity coefficients [33]. As yet another powerful example, the perturbation bounds discussed herein (particularly, Equation (9)) inspired the development of a perturbation theory for general state-space, discrete-time Markov chains [52].

In recent years, that theory has blossomed (see, e.g., Refs. [55–59]) and deserves a separate, detailed review (which, in fact, is about to be published in the context of Markov chain Monte Carlo methodology [60]). It should also be mentioned that perturbation bounds for the stationary distribution of finite state-space, discrete-time Markov chains form a now-classic topic in matrix analysis, which has been characterized by outstanding mutual enrichment of linear algebra and applied probability [40].

Moreover, a theory has been developed that utilizes Markov chain perturbation bounds as straightforward plug-ins to readily obtain sensitivity bounds for hidden Markov models [61]. Continuing the topic of Markov processes, we should mention finite-time perturbation bounds for diffusions [62]. Perturbation bounds for the stationary distributions of diffusions have also been obtained [63,64]; interestingly enough, they do not appear (unlike our Equation (10)) to be directly related to perturbation bounds that are uniform over \( t \geq 0 \). Thus, derivation and investigation of time-uniform perturbation bounds for diffusions may be a promising research topic.

Finally, an intriguing interplay between established and new results can be found in the recent work on regime-switching processes, where Markov chain considerations were used to gauge approaches to perturbation analysis for processes with a more complex structure [35,36].

A different direction of perturbation research, which, conceptually, is closely related to the material of this perspective article, has recently been developed in control theory for deterministic contractive systems. Such systems have convergence properties that can be defined using a generalization of Equation (6) (a generalization containing an expression of the form \( C\| p_1(0) - p_2(0) \| \) instead of the constant \( C \) [65,66]. Perturbation and approximation bounds for both regular [67,68] and singular [69] perturbations of contract-
tive systems have been derived. Whereas this perturbation theory developed within the domain of differential equations independently of Markov chain theory, there appear to be possibilities for cross-fertilization. Importantly, for all of these systems, the exponential rate of infinite-time convergence plays an essential role. This is a manifestation of the pattern where the parameter governing the effects of perturbations in the initial conditions also governs the effects of perturbations in the system’s parameter values, which could be suggested as a general phenomenon for dynamical systems [70]. Perhaps, in the future, such theories will converge, using the Kolmogorov equations as a shared research focus, and will continue to strengthen each other, thereby benefiting diverse applications.

These developments concern deterministic perturbations of stochastic and deterministic systems; the situation with stochastic perturbations has also been the focus of intensive research. Random perturbations of dynamical systems are a now-classic subfield of stochastic processes. Naturally for random perturbations, the rate of convergence to the unperturbed, deterministic process is typically analyzed using large-deviations theory [71,72] (as an illustration of new research, see Ref. [73]). Stochastic perturbations of stochastic systems are an area where opportunities for a relevant theory are wide open. One promising approach is to cast such a theory in the framework of stochastic processes in a random environment, where the randomness in the environment represents the perturbations, which are perhaps assumed to be small. Work in this direction has started [74,75] (including approaches based on large deviations [72]), but further progress seems to be needed before the theory is fully ready for broad applications.

7. Conclusions

The purpose of this perspective article is to illustrate the approaches and results available in the inequality-based perturbation theory for continuous-time Markov chains. Herein, our priority was to emphasize the logical interconnections between different approaches. By intention, this is not a comprehensive overview of this research field. A systematic overview should include a broader discussion of the available bounds (including the array of results centered on ergodicity coefficients for continuous-time chains [33]), a more detailed analysis of the relationships between Markov chain perturbation results in continuous and discrete time (including approaches focused on entrywise, rather than norm-based, perturbation bounds), perhaps a deeper technical dive into the proofs for the presented results, and a look into the numerical accuracy of the available perturbation bounds on practically relevant examples. It would also be informative to consider cases of unstructured perturbations of continuous-time Markov chains (i.e., cases where the perturbed system of differential equations is not a proper system of Kolmogorov equations)—a situation that can arise in numerical solution problems [1]. Each of these topics deserves a focused presentation and can motivate future studies. Contemporary perturbation theory for Markov chains is an exciting example of interdisciplinary mathematics that draws ideas and tools from probability theory, stochastic processes, differential equations, operator theory, and matrix analysis, and has the potential to impact numerous areas of applied research. We hope that this perspective article will facilitate the continued growth of this promising research direction.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The author is grateful to the Reviewers for their comments and suggestions on the manuscript.

Conflicts of Interest: The author declares no conflicts of interest.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.