The Crandall-Liggett Generation Theorem

Abstract: The Crandall-Liggett generation theorem[1] describes how to solve a broad class of differential equations by exponentiating operators. The technique is useful in both theory and practice, and illustrates the power of treating functions as points in a vector space.

The authors' proof of the paper's crucial lemma ends with the memorable sentence,

The (somewhat awkward) induction is left to the reader.

The derivation below uses a beautiful (induction-free) counting argument from elementary probability theory.

Nomenclature: The goal of the Crandall-Liggett generation theorem is to solve the initial value problem

$$\begin{cases} \dfrac{d}{dt}u(t) + A(u(t)) \ni 0, \\ \qquad\qquad u(0) = u_0 \end{cases} \qquad (IVP)$$

for a suitable class of operators $A : X \to X$. The solution is denoted $u(t) = e^{-tA}u_0$, just as if $A(\cdot)$ were the operator "multiply by a (non-negative) real number". The notation suggests — and, indeed, it is true — that

$$u(t+s) = e^{-(t+s)A}u_0 = e^{-sA}\left[e^{-tA}u_0\right]. \qquad (SG)$$

In English: The solution at time $t+s$ starting from u_0 is the solution at time s starting from $e^{-tA}u_0 = u(t)$. This is the **semi-group** property in the paper's title. The semi-group property represents the notation that physical laws do not change over time.

The operator $A(\cdot)$ is said to "generate the semi-group". The "semi-" is required because the heat equation can be solved forward in time, but not backward, so s and t are required to be non-negative.

Outline: There are 3 parts to the presentation:

1) Motivation: What class of differential equations reflect physical systems? In other words, what characteristics should $A(\cdot)$ have?

2) Method: What technique for computing e^{-tA} makes sense for the class in Part 1)? Not the power series, surely, because $A(\cdot)$ may well be unbounded, so

$$e^{-tA} = I - tA + \frac{t^2}{2!}A^2 - \frac{t^3}{3!}A^3 + \cdots$$

is surely very bad, indeed. Part 1) suggests the correct procedure.

3) Convergence: The generation theorem proves that the sequence of approximations in Part 2) is Cauchy in the Banach space X. Furthermore, the rate of convergence is part of the proof, so numerical error estimates fall out of the proof.

1) Motivation: Energy dissipates over time. The temperature difference between a cup of hot coffee and a bowl of ice cream may start out quite large, but it decreases over time. A good model for the temperature should predict this temperature contraction.

[1] M.G. Crandall and T.M. Liggett, "Generation of Semi-Groups of Nonlinear Transformations on General Banach Spaces", *Am. J. of Math*, (93), April, 1971, pp. 265–295.

Example: Newton's Law of Cooling is

$$\frac{d}{dt}u(t) + \underbrace{k(u(t) - u_a)}_{A(u)} = 0,$$

where $u(t)$ is temperature, u_a is the ambient temperature, and $k \geq 0$ is a conversion factor from temperature *gradient* to temperature *flow*. The operator $A : R \to R$ is affine but non-linear. If the temperature of the coffee is $v(t)$ and that of the ice cream is $u(t)$, then a physicist might compute (using the chain rule) that

$$\begin{aligned}\frac{d}{dt}\frac{1}{2}\Big[v(t) - u(t)\Big]^2 &= -\Big[v(t) - u(t)\Big]\Big[A(v(t)) - A(u(t))\Big] \\ &= -k\Big[v(t) - u(t)\Big]^2 \leq 0.\end{aligned}$$

In English: The square difference of the temperatures decreases over time.

Example: The previous example didn't really answer the original question: does the (absolute value) of the temperature difference decrease over time? To find out, compute

$$\begin{aligned}\frac{d}{dt}\Big|v(t) - u(t)\Big| &= -\mathrm{sgn}\,(v(t) - u(t))\Big[A(v(t)) - A(u(t))\Big] \\ &= -k\Big|v(t) - u(t)\Big| \leq 0.\end{aligned}$$

In English: The absolute difference of the temperatures decreases over time.

What's really required? If $\varphi(\cdot)$ is some measure of distance such as $|\cdot|$ or $(\cdot)^2$, then two solutions $v(t)$ and $u(t)$ get closer to each other if

$$\begin{aligned}\frac{d}{dt}\varphi\,(v(t) - u(t)) &= D\varphi(v(t) - u(t))\frac{d}{dt}\,(v(t) - u(t)) \\ &= -D\varphi(v(t) - u(t))\,(A(v(t)) - A(u(t))) \leq 0.\end{aligned}$$

Remark: The $\varphi(\cdot)$ of interest include norms and square norms, but convex analysts will point out that the theory is no more complicated for any convex function than for any special class of convex functions. If $\varphi(\cdot)$ is convex, then $D\varphi$ becomes the subdifferential $\partial\varphi$.

Remark: If $\langle\cdot,\,\cdot\rangle$ denotes the X^*, X duality pairing, then the operators A that we want satisfy

$$\langle w,\, A(v) - A(u)\rangle \geq 0 \qquad \text{for some } w \in \partial\varphi(v - u). \qquad (Acc)$$

These operators are called **accretive**. Accretiveness requires that $A(\cdot)$ be "monotonic" in some sense, but not continuous. In particular, $\partial\varphi$ may be multi-valued, which suggests $A(\cdot)$ should be permitted to be multi-valued as well. Accretive operators are therefore defined in terms of their graphs:

$$A \subset X \times X \text{ is accretive} \iff \forall (x_i, y_i) \in A,\ \exists w \in \partial\|x_2 - x_1\|_X \text{ such that } \langle w, y_2 - y_1\rangle \geq 0.$$

As far as I know, accretiveness only applies to norms, not to general $\varphi(\cdot)$.

1-d Summary: Let $A : R \to 2^R$ be a monotonic (non-decreasing) graph and let φ be a convex function on R. Then solutions to the initial value problem (IVP) exhibit the contraction property

$$\frac{d}{dt}\varphi\,(v(t) - u(t)) \leq 0.$$

Multi-dimensional Examples: Dynamics in one dimension, while illuminating, are too simple.

Example: A mass on a spring or an "LRC" electronics circuit satisfies

$$m\frac{d^2}{dt^2}u(t) + c\frac{d}{dt}u(t) + ku(t) = 0,$$

where a and b are non-negative. Physicists multiply by $\frac{d}{dt}u(t)$ and use the chain rule twice:

$$\begin{aligned} 0 &= m\frac{d}{dt}u(t)\frac{d^2}{dt^2}u(t) + c\left(\frac{d}{dt}u(t)\right)^2 + ku(t)\frac{d}{dt}u(t) \\ &= \frac{d}{dt}\left[\frac{1}{2}m\left(\frac{d}{dt}u(t)\right)^2 + \frac{1}{2}k\left(u(t)\right)^2\right] + c\left(\frac{d}{dt}u(t)\right)^2 \end{aligned}$$

The first term in square brackets is the kinetic energy, the second is the potential energy. The equation says the kinetic plus potential energy is non-increasing — just the kind of contraction we're seeking.

To cast the problem as an (IVP), convert the second-order equation into a first order equation (as always):

$$\frac{d}{dt}\begin{bmatrix} u \\ v \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & -1 \\ \frac{k}{m} & \frac{c}{m} \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix}}_{A\left(\begin{bmatrix} u \\ v \end{bmatrix}\right)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Take the dot product with $[\,ku \quad mv\,]^T$ and use the chain rule:

$$\begin{aligned} 0 &= [\,ku \quad mv\,]\frac{d}{dt}\begin{bmatrix} u(t) \\ v(t) \end{bmatrix} + [\,ku \quad mv\,]\begin{bmatrix} 0 & -1 \\ \frac{k}{m} & \frac{c}{m} \end{bmatrix}\begin{bmatrix} u \\ v \end{bmatrix} \\ &= \frac{d}{dt}\left[\frac{1}{2}ku(t)^2 + \frac{1}{2}mv(t)^2\right] + cv^2(t). \end{aligned}$$

Since $cv^2(t) \geq 0$, the derivative must be ≤ 0:

$$\frac{d}{dt}\left[\frac{1}{2}ku(t)^2 + \frac{1}{2}mv(t)^2\right] \leq 0.$$

In English: The energy is non-increasing. The "right" Hilbert space uses the weighted inner product

$$\left(\begin{bmatrix} u \\ v \end{bmatrix}, \begin{bmatrix} w \\ x \end{bmatrix}\right) = [\,u \quad v\,]\begin{bmatrix} k & 0 \\ 0 & m \end{bmatrix}\begin{bmatrix} w \\ x \end{bmatrix}.$$

In this Hilbert space, the operator A is "positive" in the sense that

$$\left(\begin{bmatrix} u \\ v \end{bmatrix}, A\left(\begin{bmatrix} u \\ v \end{bmatrix}\right)\right) = cv^2 \geq 0.$$

In fact, the frictional term modeled above by cv can be any monotonic function of v — a common example is the quadratic model $cv|v|$.

Remark: The attitude is that the Crandall-Liggett generation theorem reduces solving the initial value problem to finding the "right" space in which solutions contract over time.

Infinite dimensional Examples: The easiest example in infinite dimensions is:

Example: Let $X = \ell_\infty$, the space of bounded, real sequences. Let

$$A(\{u_1,\, u_2,\, \cdots, u_k,\, \cdots\}) = \{u_1,\, 2u_2,\, \cdots,\, ku_k,\, \cdots\}$$

Then A is linear and accretive, but unbounded. Nevertheless, the solution to (IVP) is (or, at least, should be)

$$e^{-tA}\{u_1,\, u_2,\, \cdots,\, u_k,\, \cdots\} = \{e^{-t}u_1,\, e^{-2t}u_2,\, \cdots,\, e^{-kt}u_k,\, \cdots\},$$

which is certainly in ℓ_∞ for all $t \geq 0$.

Remark: The "solution" may not be differentiable into ℓ_∞, but it *is* differentiable into X if $A(u_0) \in \ell_\infty$.

Example: The heat equation (equivalently, the diffusion equation),

$$\begin{cases} \dfrac{\partial}{\partial t}u(x,t) - \nabla_x \cdot (k(x)\nabla_x u(x,t)) = f(x), & x \in \Omega, t \geq 0, \\ u(x,t) = g(x), & x \in \partial\Omega, t \geq 0, \\ u(x,0) = u_0(x) & \end{cases}$$

exhibits "decay over time", too. As long as the "heat source" $f(x)$ and the boundary term $g(x)$ are the same for two solutions u and v,

$$\begin{aligned} \frac{d}{dt}\frac{1}{2}\left\|v(\cdot,t) - u(\cdot,t)\right\|^2_{L^2(\Omega)} &= \frac{d}{dt}\frac{1}{2}\int_\Omega \Big(v(x,t) - u(x,t)\Big)^2\, dx \\ &= \int_\Omega \Big(v(x,t) - u(x,t)\Big)\, \nabla_x \cdot \Big(k(x)\nabla_x\left(v(x,t) - u(x,t)\right)\Big)\, dx \\ &= -\int_\Omega k(x)\left\|\nabla_x v(x,t) - \nabla_x u(x,t)\right\|^2_{R^n}\, dx \leq 0, \end{aligned}$$

where the last manipulation is integration by parts (the equal boundary values of u and v make the boundary integral zero).

Example: The thermal properties of an object may be temperature-dependent. For example, the diffusivity (k in the equation above) of ice is different from the diffusivity of water or steam. The phase change is modeled by

$$\frac{d}{dt}u(t) - \nabla\cdot\nabla\alpha(u) = 0,$$

where $\alpha(x) = \int^x k(t)\,dt$ is monotone (non-decreasing) because the integrand $k \geq 0$. If v and u are solutions to (IVP), if φ is convex, and if $\langle\cdot,\cdot\rangle$ is the duality paring on some $L^p(\Omega)$, then integration by parts gives

$$\begin{aligned} 0 &= \left\langle \partial\varphi(v-u),\, \frac{d}{dt}(v-u)\right\rangle - \Big\langle \partial\varphi(v-u),\, \nabla\cdot\nabla\left(\alpha(v) - \alpha(u)\right)\Big\rangle \\ &= \frac{d}{dt}\varphi(\beta(v) - \beta(u)) + \int_\Omega \nabla\partial\varphi(v-u)\cdot\nabla(\alpha(v) - \alpha(u))\, dx \end{aligned}$$

The goal is to make the rightmost integrand non-negative. If $\partial\varphi(v-u) = \beta\left(\alpha(v) - \alpha(u)\right)$ for some monotonic (non-decreasing) function $\beta(\cdot)$, then

$$\nabla\partial\varphi(v-u)\cdot\nabla(\alpha(v) - \alpha(u)) = \beta'(\alpha(v) - \alpha(u))\left\|\nabla(\alpha(v) - \alpha(u))\right\|^2 \geq 0.$$

One easy choice (modulo differentiability questions) is $\partial\varphi(x) = \mathrm{sgn}\,(x)$ because, by the monotonicity of $\alpha(\cdot)$, $\mathrm{sgn}\,(v-u) = \mathrm{sgn}\,(\alpha(v) - \alpha(u))$. Consequently, heat transfer with a phase change is modeled in $X = L^1$.

2) Method: What technique successfully computes $u(t) = e^{-tA}u_0$ for (possibly non-linear, unbounded) accretive operators $A(\cdot)$? The power series

$$e^{-tA} = I - tA + \frac{t^2}{2!}A^2 - \frac{t^3}{3!}A^3 + \cdots.$$

can not be successful with unbounded operators. Instead, replace the differential equation with a difference equation:

$$u(t+h) - u(t) \in -hA(u(?)).$$

What should the "?" be? If "?" $= t$, then

$$\begin{aligned} u(h) &\in u_0 - hA(u_0) \\ u(2h) &\in u(h) - hA(u(h)) \\ &\in u_0 - hA(u_0) - hA\left(u_0 - hA(u_0)\right) \\ &\vdots \end{aligned}$$

This procedure, called **Euler's explicit method**, obviously fails when A is unbounded. If, however, "?" $=$ $t+h$, then the procedure, called **Euler's implicit method**, is

$$u(t+h) + hA(u(t+h)) \ni u(t).$$

In other words,

$$u(t+h) = \underbrace{(I + hA)^{-1}}_{J_h} u(t),$$

assuming the operator J_h is well-behaved. In particular, Euler's implicit approximation is, then,

$$\begin{aligned} u(h) &= \Big(I + hA\Big)^{-1} u_0 \\ u(2h) &= \Big(I + hA\Big)^{-2} u_0 \\ &\vdots \\ u(nh) &= \Big(I + hA\Big)^{-n} u_0. \end{aligned}$$

The question boils down to the behavior of the **resolvent**

$$J_h = \Big(I + hA\Big)^{-1}. \qquad (Res)$$

Lemma: If $A(\cdot)$ is accretive and $h > 0$, then J_h is a contraction.

Proof: If $(x_i, y_i), i = 1, 2$ are in the graph of A, then there is a $w \in \partial\|x_2 - x_1\|$ such that $\langle w, y_2 - y_1\rangle \geq 0$. Since $\|w\|_{X^*} \leq 1$,

$$\begin{aligned} \|x_2 - x_1\| &= \langle w, x_2 - x_1\rangle \\ &\leq \langle w, (x_2 + hy_2) - (x_1 + hy_1)\rangle \\ &\leq \|(x_2 + hy_2) - (x_1 + hy_1)\| \end{aligned}$$

The lemma follows from the observation that $x_i + hy_i \in (I + hA)(x_i)$, so $x_i \in (I + hA)^{-1}(x_i + hy_i)$. ////

Remark: The lemma is not unexpected: the operator e^{-tA} is a contraction, and a good approximation should also be a contraction.

Remark: The negative Laplacian, $-\Delta$, is an unbounded operator on lots of function spaces, but the resolvent $(I - h\Delta)^{-1}$ is (hypo-)elliptic — very well behaved, indeed! We may therefore expect to solve the heat equation (a parabolic equation) by approximating with the solutions to a sequence of elliptic equations.

Remark: The domain of $A(\cdot)$ may be "small", but the range of $I + hA$ (and, therefore, the domain of J_h) may very well be all of X. Accretive operators for which $R(I + hA) = X$ are called **m-accretive**.

Theorem: The Crandall-Liggett Generation Theorem: If A is m-accretive, then for any $T > 0$, the sequence

$$u_n(T) = \left(I + \frac{T}{n}A\right)^{-n} u_0 = J_{\frac{T}{n}}^n u_0$$

converges for every $u_0 \in D(A)$. Furthermore, the rate of convergence is at least sublinear, with

$$\|u_n(T) - u_m(T)\| \leq 2T\sqrt{\left|\frac{1}{m} - \frac{1}{n}\right|} \, \|A(u_0)\| \, .$$

Remark: Since e^{-tA} is a contraction, the sequence $u_n(T)$ converges for every $u_0 \in \overline{D(A)}$.

The proof consists of "peeling off" contractions $J_{\frac{T}{n}}$. To do so, we use the following lemma:

Lemma: If $0 < h < k$ and A is m-accretive, then $y = J_k(x)$ iff

$$\begin{aligned} x \in y + kA(y) \quad &\iff \\ \frac{h}{k}x \in \frac{h}{k}y + hA(y) \quad &\iff \\ \frac{h}{k}x + \frac{k-h}{k}y \in y + hA(y) \quad &\iff \\ J_h\left(\frac{h}{k}x + \frac{k-h}{k}J_k(x)\right) = y = J_k(x). & \end{aligned}$$

The bottom line is called the **resolvent formula**.

Remark: The argument of J_h above is

$$\frac{h}{k}x + \frac{k-h}{k}J_k(x) = J_k(x) + h\left(\frac{I - J_k(x)}{k}\right),$$

so the lemma implies

$$J_k(x) + h\left(\frac{I - J_k(x)}{k}\right) \in (I + hA)(J_k(x)).$$

Consequently,

$$A_k(x) \equiv \left(\frac{I - J_k(x)}{k}\right) \in A(J_k(x)).$$

A_k is called a **Yoshida approximation** of A.

Example: Let $A(x) = \operatorname{sgn}(x)$. Then

$$J_h(x) = (I + h\operatorname{sgn}(\cdot))^{-1}(x) = \begin{cases} x - h, & x > h \\ 0, & -h \leq x \leq h \\ x + h, & x < -h \end{cases}$$

so

$$A_h(x) = \frac{1}{h}(I - J_h) = \begin{cases} 1, & x > h \\ \frac{x}{h}, & -h \leq x \leq h \\ -1, & x < -h \end{cases}$$

Note in particular that A_h is a Lipschitz continuous under-estimator of A (in the sense that $|A_h(x)| \le |y|$ for all $y \in A(x)$).

Proof: (of the CLGT.) Let $0 < m < n$ be integers, and let $\frac{T}{n} = h < \frac{T}{m} = k$. Then the resolvent formula, the contractiveness of J_h, and the convexity of the norm imply

$$\begin{aligned}
\|J_h^n(u_0) - J_k^m(u_0)\| &= \left\| J_h^n(u_0) - J_h\left(\frac{h}{k} J_k^{m-1}(u_0) + \frac{k-h}{k} J_k^m(u_0)\right)\right\| \\
&\le \frac{h}{k}\left\|J_h^{n-1}(u_0) - J_k^{m-1}(u_0)\right\| + \frac{k-h}{k}\left\|J_h^{n-1}(u_0) - J_k^m(u_0)\right\|
\end{aligned}$$

Let $p = \frac{h}{k} = \frac{m}{n}$ and $q = 1 - p = \frac{k-h}{k} = \frac{n-m}{n}$. Iterating on the estimate above,

$$\begin{aligned}
\left\|J_h^n(u_0) - J_k^m(u_0)\right\| &\le p\left\|J_h^{n-1}(u_0) - J_k^{m-1}(u_0)\right\| + q\left\|J_h^{n-1}(u_0) - J_k^m(u_0)\right\| \\
&\le p^2\left\|J_h^{n-2}(u_0) - J_k^{m-2}(u_0)\right\| + 2pq\left\|J_h^{n-2}(u_0) - J_k^{m-1}(u_0)\right\| \\
&\quad + q^2\left\|J_h^{n-2}(u_0) - J_k^m(u_0)\right\| \\
&\le p^3\left\|J_h^{n-2}(u_0) - J_k^{m-3}(u_0)\right\| + 3p^2q\left\|J_h^{n-3}(u_0) - J_k^{m-2}(u_0)\right\| \\
&\quad + 3pq^2\left\|J_h^{n-3}(u_0) - J_k^{m-1}(u_0)\right\| + q^2\left\|J_h^{n-3}(u_0) - J_k^m(u_0)\right\|
\end{aligned}$$

Remark: Note the similarity to the binomial expansion.

Remark: Note that the expansion holds for any convex function, not just $\|\cdot\|$ (provided J_h is a contraction relative to the convex function, of course).

To "see" the expansion, construct an $n \times m$ lattice, with lattice point (i, j) representing the term of the form

$$\gamma_{i,j}\left\|J_h^i(u_0) - J_k^j(u_0)\right\|.$$

The expansion stops when $j = 0$ (the left edge) or $i = 0$ (the bottom edge). To save space, omit the norm and record only the coefficient $\gamma_{i,j}$:

						1
					p	q
				p^2	$2pq$	q^2
			p^3	$3p^2q$	$3pq^2$	q^3
		$\iddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	p^{m-1}	$\cdots$	$\binom{m-1}{3}p^3q^{m-4}$	$\binom{m-1}{2}p^2q^{m-3}$	$\binom{m-1}{1}pq^{m-2}$	q^{m-1}
p^m	$\binom{m}{m-1}p^{m-1}q$	$\cdots$	$\binom{m}{3}p^3q^{m-3}$	$\binom{m}{2}p^2q^{m-2}$	$\binom{m}{1}pq^{m-1}$	q^m
$\binom{m}{m-1}p^mq$	$\binom{m+1}{m-1}p^{m-1}q$	$\cdots$	$\binom{m+1}{3}p^3q^{m-2}$	$\binom{m+1}{2}p^2q^{m-1}$	$\binom{m+1}{1}pq^m$	q^{m+1}
$\vdots$	$\vdots$	$\iddots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\binom{n-2}{m-1}p^mq^{n-m-1}$	$\binom{n-1}{m-1}p^{m-1}q^{n-m}$	$\cdots$	$\binom{n-1}{3}p^3q^{n-4}$	$\binom{n-1}{2}p^2q^{n-3}$	$\binom{n-1}{1}pq^{n-2}$	q^{n-1}
$\binom{n-1}{m-1}p^mq^{n-m}$	$\binom{n}{m-1}p^{m-1}q^{n-m+1}$	$\cdots$	$\binom{n}{3}p^3q^{n-3}$	$\binom{n}{2}p^2q^{n-2}$	$\binom{n}{1}pq^{n-1}$	q^n

(The empty spaces in the upper left triangle are zeros.)
The counting logic is different for the bottom and left edges (where the expansion must stop). The bottom edge is easier to analyze counting from right to left. To reach the bottom in position $j = m - \ell$, the path must take n legs total, ℓ of which must be in the direction $\swarrow$. There are $\binom{n}{\ell}$ such legs, so the coefficient is

$$\gamma_{m-\ell,\,0} = \binom{n}{\ell} p^{\ell} q^{n-\ell}, \qquad 0 \le \ell \le m-1.$$

Consequently, the bottom edge (excluding the corner element) contributes

$$\sum_{\ell=0}^{m-1} \binom{n}{\ell} p^{\ell} q^{n-\ell} \left\| u_0 - J_k^{m-\ell}(u_0) \right\|$$

to the error.

Likewise, to reach the left edge at position $i = n - \ell$, the path must take ℓ legs. The path must pass through $(1, n-\ell-1)$ because the only paths to the left edge must move in the direction $\swarrow$ on the last leg. There are $\binom{\ell-1}{m-1}$ such legs, so the coefficient is

$$\gamma_{0,\,n-\ell} = \binom{\ell-1}{m-1} p^{m} q^{\ell-m}, \qquad m \le \ell \le n.$$

Consequently, the left edge (including the corner element) contributes

$$\sum_{\ell=m}^{n} \binom{\ell-1}{m-1} p^{m} q^{\ell-m} \left\| J_h^{n-\ell}(u_0) - u_0 \right\|$$

to the error.

Combining the two edges gives the estimate

$$\|J_h^n(u_0) - J_k^m(u_0)\| \le \sum_{\ell=0}^{m-1} \binom{n}{\ell} p^{\ell} q^{n-\ell} \left\| u_0 - J_k^{m-\ell}(u_0) \right\| + \sum_{\ell=m}^{n} \binom{\ell-1}{m-1} p^{m} q^{\ell-m} \left\| J_h^{n-\ell}(u_0) - u_0 \right\|. \qquad (E_0)$$

Remark: The bottom row's distribution is part of the **Binomial distribution**. The left column's distribution is part of the **Negative Binomial distribution**.

The proof now follows from two lemmas (with 2 estimates each):

Lemma: For all $h \ge 0$, for all $x \in D(A)$, and for any $y \in A(x)$:
1) $\left\| J_h(x) - x \right\| \le h \left\| y \right\|$. Informally, $\left\| J_h(x) - x \right\| \le h \left\| A(x) \right\|$.
2) $\left\| J_h^{\ell}(x) - x \right\| \le \ell h \left\| y \right\|$.

Proof: 1) Suppose $(x, y) \in A$. Then $x = J_h(x + hy)$. Since J_h is a contraction,

$$\left\| hy \right\| \ge \left\| J_h(x+hy) - J_h(x) \right\| = \left\| x - J_h(x) \right\|.$$

Part 1) follows by dividing by h.

Part 2) follows from part 1) by making the sum telescope:

$$\left\| J_h^{\ell}(x) - x \right\| = \left\| \sum_{i=1}^{\ell} J_h^{i}(x) - J_h^{i-1}(x) \right\| \le \sum_{i=1}^{\ell} \left\| J_h^{i}(x) - J_h^{i-1}(x) \right\| \le \sum_{i=1}^{\ell} \left\| J_h(x) - x \right\| \le h\ell \left\| A(x) \right\|.$$

////

The lemma turns estimate (E_0) into

$$\|J_h^n(u_0) - J_k^m(u_0)\| \leq \left[\sum_{\ell=0}^{m-1} k(m-\ell)\binom{n}{\ell} p^\ell q^{n-\ell} + \sum_{\ell=m}^{n}(n-\ell)h\binom{\ell-1}{m-1}p^m q^{\ell-m}\right]\left\|A(u_0)\right\|. \qquad (E_1)$$

The first sum is an expectation involving a binomial random variable $Y \sim Bin(n,p)$; the second is an expectation of a negative binomial random variable $Z \sim NB(m,p)$. The mean and variance of Y and Z are well-known (among probabilists and statisticians):

$$\begin{aligned} E(Y) &= np = m \\ Var(Y) &= npq \\ E(Z) &= \frac{m}{p} = n \\ Var(Z) &= \frac{mq}{p^2} = \frac{n(n-m)}{m} \end{aligned}$$

Armed with these values, Chebyshev's inequality applied to the two sums is

$$\begin{aligned} \sum_{\ell=0}^{m-1} k(m-\ell)\binom{n}{\ell}p^\ell q^{n-\ell} &= kE(E(Y)-Y)^+ \leq k\sqrt{Var(Y)} \\ &\leq k\sqrt{npq} = \frac{T}{m}\sqrt{n\frac{m}{n}\frac{n-m}{n}} = T\sqrt{\frac{1}{m}-\frac{1}{n}}, \end{aligned}$$

and

$$\begin{aligned} \sum_{\ell=m}^{n}(n-\ell)h\binom{\ell-1}{m-1}p^m q^{\ell-m} &= hE(E(Z)-Z)^+ \leq h\sqrt{Var(Z)} \\ &\leq \frac{T}{n}\sqrt{\frac{n(n-m)}{m}} = T\sqrt{\frac{1}{m}-\frac{1}{n}}, \end{aligned}$$

The two estimates together turn estimate (E_1) into

$$\|J_h^n(u_0) - J_k^m(u_0)\| \leq 2T\sqrt{\frac{1}{m}-\frac{1}{n}}. \qquad (E_2)$$

Consequently, the sequence $\left(I + \frac{T}{n}A\right)^{-n} u_0$ is Cauchy in X, and therefore converges. ////