Magnetic Analysis by a Flux-Centered Approach
Part 2: Basic Equations for Solenoids with Ideal Ferromagnetic Conductor

Joseph B. Seale

This paper is written to stand on its own and therefore repeats some of the formulation found in Part 1. The paper summarizes basic, known physics at it applies specifically to solenoids, ignoring non-ideal ferromagnetic behavior.

Governing Equations

No more mechanical energy can be recovered from a solenoid actuation stroke than goes in, electrically. At any fixed armature position, a solenoid looks like an inductor, capable of storing energy in its internal magnetic field. When the armature moves, changing the field geometry, some fraction of the field energy is converted to or from mechanical energy. This conversion can be characterized, starting with the energy equation for a linear inductor:

1] $E = ½ I^2L$ energy stored in linear inductor

It is recognized that solenoids are regularly pushed into partial saturation, where inductance L becomes a variable function of current I and the energy integration leading to Eq. 1 is invalidated. Maintaining magnetic materials in saturation requires a sustained application of magnetomotive force, i.e. ampere-turns in a winding, which implies energy expenditure. Efficient solenoids simply cannot operate with substantial fractions of their material in deep saturation. For practical purposes, Eq. 1 can be taken as at least approximately true of solenoids in their efficient realm of operation. This leads us to the fundamental relationship of energy to mechanical force. Given an increment of change, dX, in the position of an armature, driven by electromagnetic force, F, then work, dW, is done on the armature, creating a counterbalancing change in electromagnetic energy, -dE, as expressed by Eq. 2:

2] $– dE = dW = FdX$ increment of work energy exerted at force F through motion dX

It is convenient to define position X in terms of the gap between an armature and yoke, such that X = 0 represents the magnetic closed position of armature and yoke, while X>0 for other armature positions. If force is defined positive as pushing the armature to increasing X, consistent with the sign conventions of Eq. 2, then force of magnetic attraction can be expected to be negative, i.e. F < 0. Dividing the energy increment –dE by the distance increment dX gives a derivative expression carried from Eq. 2 to Eq. 1:

3] $F = -dE/dX = – d(½ I^2L)/dX$ force matches the change in magnetic energy with X

In performing the differentiation on the right of Eq. 3, it must be recognized that both inductance L and current I vary as functions of X. Eq. 3 is valid only if the energy transfer associated with force F involves only a conversion of magnetic energy in the solenoid, with no additional energy flowing in and out through the windings while X changes. Imagining idealized conditions under which no energy transfer to or from winding wires would take place, consider a hypothetical winding of superconductive material, connected to a power source to energize the solenoid at constant armature position X, then shorted so that current continues to flow but no energy transfer takes place. To characterize this situation, it is useful to introduce the flux linkage, $n\Phi$ (commonly represented as λ), for magnetic flux Φ effectively linking n turns of winding. In a solenoid whose magnetic geometry changes considerably with changing armature position X, the distribution of flux through the windings may change, with some flux failing to link some of the n winding turns. In this instance, the definition of flux Φ may be ambiguous: how much flux counts? For the purposes of exchanging energy between an electrical source and a winding, flux only counts if it links turns of wire. The expression "$n\Phi$" represents the integral of flux over all the turns in the winding. Given n turns of wire, the flux that "counts" is given by $\Phi = (n\Phi)/n$. The expression $n\Phi$ is further defined by the inductive voltage Vi:

4] $Vi = d(n\Phi)/dt$ inductive voltage equals time derivative of flux linkage, often called λ

Like Eq. 1, Eq. 4 is not derived here. It comes straight out of fundamental physics. Under conditions of no energy exchange with an external power source, and no energy loss to winding resistance, we have the special situation defined by Eq. 5:

5] $Vi = 0 = d(n\Phi)/dt$ for shorted superconductive winding

Inductance L may be related to flux linkage $n\Phi$, since inductance defines the ratio of inductive voltage Vi to rate of change of current, dI/dt:

6] $Vi = L\, \partial I/\partial t\big|_L = d(n\Phi)/dt$ L relates I to $n\Phi$ by partial derivative at constant L

It may not be clear what Vi should be in the case where inductance L is varying over time, as in a moving solenoid. Eq. 6 states that under partial derivative conditions of current with inductance held constant, we can unambiguously equate inductance multiplied by the partial time derivative of current with the time derivative of flux linkage. Integrating Eq. 6 over time as current I changes at constant inductance, L, yields:

7] $IL = n\Phi\ (= \lambda)$ current times inductance equals flux linkage

Although derived for the special case where L is held constant, Eq. 7 expresses an equation of state, whose validity holds regardless of the path taken to reach that state. If a flux linkage $n\Phi$ is reached by some path of varying L and varying I, the result is always the same when a final inductance L is reached, and current I must be related to L and Φ in the manner stated. In a 2-dimensional plane of inductance L and current I, Eq. 6 represents one particular path for reaching the final state of Eq. 7. Considering other paths in the (I,L) plane, Eq. 7 leads us to a more general definition of inductance under conditions of time-varying geometry. Eq. 5 has general physical validity, grounded in Maxwell's equations. When flux linkage changes, whether because of current change or geometry change, then the inductive voltage Vi is given by the time derivative of $n\Phi$, as in Eq. 5. Combining Eqs. 5 and 7 leads to a more general definition of L:

8] $V_i = d(IL)/dt = L\, \partial I/\partial t\,|_L + I\, \partial L/\partial t\,|_I$ total derivative of IL as partials at constant L and I

In the special case of a shorted winding as expressed in Eq. 5, Vi = 0, leading to:

9] $0 = d(IL)/dt$ for shorted superconductive winding

10] $IL = \text{constant}$ for shorted superconductive winding

11] $d(IL)/dX = 0$ for shorted superconductive winding

Eq. 9 follows from Eq. 8 where a shorted superconductive winding physically constrains inductive voltage to be zero. Eq. 10 follows from time-integration of Eq. 9. Eq. 11 appears to follow from Eq. 10, but more rigorously, Eq. 11 flows from Eq. 7. It is not possible to alter the flux linkage, $n\Phi$, in a shorted superconductive winding, because any infinitesimal change in flux linkage would induce a voltage, which cannot exist in a shorted winding of zero resistance except under transient conditions, as magnetic fields equilibrate, or under conditions of electromagnetic wave propagation from the winding. Considering inductance at low frequencies, where electromagnetic wave emission is negligible, $n\Phi$ cannot be altered by moving a solenoid armature to a different position, X, since any infinitesimal change in flux linkage would induce a change in current, I, that would have the effect of opposing the change in flux linkage. It is by this principle that a superconductive disk or ring, when lowered over a permanent magnet, will hover indefinitely over the magnet, the induced circulating current opposing net flow of magnetic flux through any closed loop of superconductive material. The induced current interacts with the permanent magnetic field to produce levitation. Eq. 12 follows from Eq. 11:

12] $I\, \partial L/\partial X + L\, \partial I/\partial X = 0$ for shorted superconductive winding

Eqs. 11 and 12 provide a means for rewriting Eq. 3 in a more revealing form:

13] $F = -\tfrac{1}{2}\left(\partial I/\partial X \bullet IL + I \bullet \partial(IL)/\partial X\right) = -\tfrac{1}{2}\left(\partial I/\partial X \bullet IL\right)$

The second expression for F on the far right of Eq. 13 follows since Eq. 11 assures that the $\partial(IL)/\partial X$ term is zero. Rearranging the far right of Eq. 13 as "$(L\partial I/\partial X) \bullet I$" leads to a substitution from Eq. 12:

14] $F = \tfrac{1}{2} I^2\, \partial L/\partial X$

If we had considered the I^2 term on the right of Eq. 3 to be constant and carried the differentiation d/dX inside to the inductance term, the result would have been:

~~14]~~ $F = -\tfrac{1}{2} I^2\, \partial L/\partial X$ <u>incorrect</u> force inference from Eq. 3, ignoring $\partial I/\partial X$

Under the physical conditions assumed for deriving electromagnetic force, the sum of magnetic and mechanical energy was taken to be a conserved, constant quantity. To express this, we considered a hypothetical shorted superconductive winding to prevent any dissipation of electrical energy or energy exchange with an external source or sink. Thus, current I was allowed to change freely as a function of armature position X. It was observed that current would change in a manner to maintain a constant flux linkage, $n\Phi$, in the hypothetical superconductive winding. It was found that a reasonable generalization of the meaning of inductance to conditions of variable geometry leads to the equality of the "inductive momentum" expression IL to the flux linkage expression $n\Phi$. When an armature is moved from a closed position with a solenoid yoke, opening up a gap in the flux path, the effect is generally to reduce inductance L, so that $\partial L/\partial X < 0$ under most circumstances in solenoids. Peculiar geometries could cause a positive $\partial L/\partial X$ where one flux-conducting surface moves past another as an armature moves "outward" from a closed position, but for simple geometries, $\partial L/\partial X$ is always negative. The change in magnetic energy as an armature is pulled away from a yoke (i.e., increasing X) due to change of inductance at constant current is negative: less inductance, less magnetic energy. As the armature is pulled away, however, the resulting increase in current I increases the magnetic energy, by exactly twice the amount by which the decline in inductance tended to reduce the magnetic energy. Thus, the sign in Eq. 14 is positive, which with a negative derivative $\partial L/\partial X$ leads to a negative force, i.e. a force of attraction, tending to reduce X and close the magnetic gap.

Aside from deriving a correct expression, the importance of the relationships expressed above lies in an understanding of how solenoid current tends to change when the armature is moved. When one substitutes a real electrical winding, with resistance and with connection to a driver circuit, for the shorted superconductive winding assumed for the physics derivation above, one arrives at the more complex situation where one can add and remove solenoid energy via the winding. Eq. 14 remains a valid general expression for force, and the derivation leads one to expect that when a solenoid armature is pulled out, electrical current will tend to increase. When an armature pulls itself closed, electrical current is driven down, at least until magnetic flux reaches a saturation limit, so that an increase in flux linkage can no longer produce an inductive voltage to oppose further increases in current. With a steady applied voltage, then, current rises, a solenoid is driven to close, closure drives the current back down, and finally current rises again rapidly as further increases in magnetic flux are limited by core saturation.

From Eq. 14 we can derive an expression for electromechanical work, dW, performed when a solenoid current flows and the armature moves through distance dX:

15] $$dW = F dX = \tfrac{1}{2} I^2 dL$$

In pulling a solenoid closed, the force F is negative or attractive, the driven change in position dX is negative, or toward closure, and the work dW is positive. In the case of an automotive valve solenoid, in the absence of gas pressure interaction, a quick magnetic release from latching the armature on one side will allow a spring to inject kinetic energy into the valve and armature, and this kinetic energy will be approximately enough to bring the armature to its magnetically closed position on the opposite side, with little electrical work needed. When gas pressure opposes the opening of a valve, energy is lost to the opposing gas flow, and positive increments dW of work must be supplied to overcome the loss, bringing the armature to magnetic closure. The question arises, what pattern of dW is most efficient, and what pattern of dW transfers the most possible energy, given saturation constraints. Inductance L increases as X decreases with a closing solenoid, so increments dL are multiplied by a positive current-squared to produce increments of work. The resistive energy loss Er associated with current is current-squared times resistance:

16] $dEr/dt = I^2R$

While resistive losses Er are accumulating in proportion to the square of current, increments in net work are also accumulating in proportion to the square of current, as is shown in form parallel to Eq. 16 deriving from Eq. 15:

17] $dW/dt = ½\ I^2\ dL/dt = ½\ I^2\ (\partial L/\partial X)(dX/dt)$

With this information, one can determine the derivative ratio of work performed, dW, to resistive energy lost, dEr:

18] $dW/dEr = ½(dL/dt)/R = ½(\partial L/\partial X)(dX/dt)/R$

To obtain efficient actuation, one wants to maximize the ratio dW/dEr, implying that one should apply as much electrical power as possible when inductance L is varying rapidly. As the right-hand expression of Eq. 18 implies, a rapid variation in L, i.e. a high dL/dt, depends on a large slope of inductance with position, dL/dX, combined with a high armature velocity, dX/dt. If one could change current I very rapidly to any level, and if Eq. 18 were valid at all currents, then efficiency would be maximized by waiting, in the armature trajectory, until dL/dt achieves its largest magnitude, at which point one would fire a large, short current pulse to accomplish the needed electromechanical work. Eq. 18 is only valid for an unsaturated core. As a first approximation to a saturation boundary, one can say that flux Φ cannot exceed a maximum limit, or equivalently, the flux linkage $n\Phi = IL$ is limited by some bound:

19a] $n\Phi \le n\Phi_{max}$ flux linkage saturation constraint, valid for some solenoids
19b] $IL \le (IL)_{max}$ "IL" saturation constraint, equivalent to flux linkage constraint

Force equation 14, expressed in terms of current and inductance, can be re-expressed in terms of flux and effective magnetic gap. We begin with Eq. 20, which expresses force in terms of the derivative of reciprocal inductance:

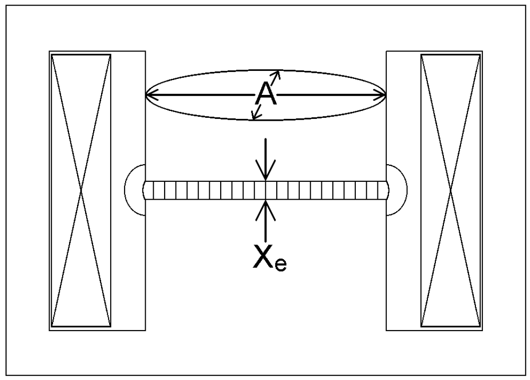
20] $F = -\ ½\ (IL)^2(\partial(1/L)/\partial X)$

To verify Eq. 20, note that expanding the partial derivative $\partial(1/L)/\partial X$ into the equal expression $(-1/L^2)\partial L/\partial X$ immediately leads to Eq. 14. Substituting from Eq. 7, the magnetic momentum product IL becomes the flux linkage $n\Phi$:

21] $F = -\ ½\ (n\Phi)^2(\partial(1/L)/\partial X)$

The reciprocal inductance function 1/L has the advantage of being a well-behaved function of X as X approaches zero. It is useful to express 1/L in terms of an effective magnetic gap, X_e. Consider the idealized physical inductor realization of Fig. 1.

Fig. 1



A toroidal magnetic circuit of an idealized, infinitely permeable material includes a narrow gap of axial width X_e and area A. The image of Fig. 1 describes a gapped pot core realistically. In the limit as X_e approaches zero, Eq. 22 correctly describes inductance L:

22] $L = \mu_o n^2 A/X_e$

If there were no fringing field and all the magnetic flux consisted of straight parallel lines out to the edges of the polefaces in Fig. 3, then Eq. 22 would be exact.

Another way to interpret Eq. 22 is to define X_e in terms of actual inductance, as expressed by rearranging Eq. 22 into Eq. 23:

22] $X_e = \mu_o n^2 A/L$

X_e is closely related to magnetic reluctance in a flux circuit. As defined by Eq. 22, with a dimension of length (meters in SI units), X_e correlates with physical magnetic gaps in solenoids in a useful way. Like reluctance, X_e represents a kind of resistance to magnetic flux, and simple formulas for series and parallel resistors can be used, with physical distances, to approximate X_e in real circuits. For a single-gap inductor like Fig. 3, or for a solenoid having reluctance dominated by a gap at the end of an armature, the path for fringing flux acts approximately like a parallel magnetic resistor, increasing the magnetic conductance represented by the reciprocal, $1/X_e$, as expressed by Eq. 23:

23] $1/X_e \approx 1/(X+X_{min}) + 1/X_o$ approximation, valid for small gaps

In Eq. 23, X is the actual geometric gap that one might find in a solenoid or gapped pot core. X_{min} represents a minimum effectively achievable gap when the face of an armature is at a limit stop or against a mating yoke poleface. Part of X_{min} comes from the less-than-infinite permeability of core material. An additional additive component comes from gaps where magnetic polefaces fail to meet parallel or flat. X_o represents an equivalent leakage "gap," referenced to the same flux area A, for fringing flux. In the case of a solenoid with a double gap, e.g., a UI or EI core where the "I" armature is drawn into contact with the "U" or "E" yoke, the flux in the magnetic circuit has to bridge two gaps per circuit, having possibly equal areas A_1 and A_2. In the case of an "E" core, A_1 might represent the area of the center leg while A_2 might represent the sum of the areas of the two side legs – since the side legs act magnetically in parallel, additively carrying the same flux as the center leg. For double-gapped cores, Eq. 23 works well if area "A" is defined as a "reduced area" according to Eq. 24:

24] $1/A = 1/A_1 + 1/A_2$ "reduced area" formula for double gap

In double-gap solenoids, as with E-I and U-I and pot core topologies, the winding or windings do not enclose the opening gap or gaps, so flux readily strays through lateral paths that fail to reach out and include the armature. This stray flux accounts for the X_o term in Eq. 23. In the case of conventional long cylindrical-armature solenoids with a single axial gap, Eq. 23 may be a poor approximation. In such solenoids, the cylindrical armature is attracted to close an axial flux gap at one end. The fixed return flux path goes radially across an annulus where the armature slides through a bushing in the yoke. Armature travel commonly exceeds armature diameter. The winding surrounds the axial armature gap, tending to confine the magnetic field to a straight cylindrical path through the gap, thus maintaining some significant field strength even for large gaps. In these geometries, the effective gap X_e is often observed to vary in almost linear proportion to the actual gap X. The X_{min} term is often a substantial fraction of total armature travel, including two major components: padding where the end of the armature closes with the yoke; and the annular gap where the armature slides through a bushing in the yoke. This linear fit is expressed by Eq. 25:

25] $X_e = X + X_{min}$ approximation for some long cylindrical armature solenoids

If a poleface is angled or conically tapered at an angle θ away from flat, and if A_x represents the a<u>x</u>ial projection of area, then the area of adjacent surfaces is increased by the factor $1/\cos(\theta)$ while the gap normal to the inclined surfaces is reduced, relative to the axial gap, by the factor $\cos(\theta)$. If X represents axial travel rather than magnetic gap normal to gap-forming surfaces of armature and yoke, then the factor $1/\cos(\theta)$ should be incorporated twice in a correction formula to get from axially-projected area to a corrected area measure appropriate for use in Eq. 22, as indicated by Eq. 26:

26] $A = A_x/\cos^2(\theta)$ formula area as function of angle and axially projected area

Eq. 26 is applicable for cylindrical armatures with conical ends as well as variations on E-I and U-I solenoids where the air gaps are sloping rather than square to the axial travel.

Having described the effective gap function X_e in terms of actual geometric variables and empirical function fits, we are in a position to substitute from Eq. 22 into Eq. 21 to define magnetic force F in terms of flux linkage $n\Phi$ and effective gap X_e. A few intermediate results are worth showing, first of all, an expression for reciprocal inductance:

27] $1/L = X_e/(\mu_o n^2 A)$

The next definition may look too obvious, but it is not. Flux linkage $n\Phi$ has a well defined physical meaning: its variation with time gives the voltage Vi induced in a coil, according to Eq. 4. The definition of flux Φ is not obvious – flux where in a circuit? It is useful to define an average effective flux in a solenoid as the flux linkage divided by the number of turns in a winding, n, according to Eq. 28:

28] $\Phi = (n\Phi)/n$ definition of average flux in terms of flux linkage $n\Phi$ and n windings

29] $F = (-\tfrac{1}{2}\Phi^2/\mu_o A)(\partial X_e/\partial X)$ non-saturating force in terms of flux, effective gap

For very small poleface gaps, X_e approaches X asymptotically, so the partial derivative expression in Eq. 29 approaches unity in the small-gap limit:

30] $F \approx -\tfrac{1}{2}\Phi^2/\mu_o A$ limit force expression for small magnetic gaps

Eq. 30 is also a good approximation where Eq. 25 works better than Eq. 23, e.g., for many tubular solenoids, provided that the area A is chosen well, generally for empirical optimization of fit.

In double-gapped solenoids, where Eq. 23 frequently gives a very good approximation of effective gap, the partial derivative expression of Eq. 29 is readily derived using Eq. 23, leading to Eq. 31:

31] $F = (-\tfrac{1}{2}\Phi^2/\mu_o A)/(1+(X+X_{min})/X_o))^2$ force, double-gap empirical formula

It is instructive to rewrite Eq. 31 in terms of the corrected geometric gap, $X+X_{min}$, and the effective gap, X_e:

32] $F = (-\tfrac{1}{2}\Phi^2/\mu_o A)\bullet(X_e/(X+X_{min}))^2$

From the parallel resistor analogy of Eq. 23, one can define an x-dependent armature flux, Φ_x, that flows through the “resistor” associated with reluctance $X+X_{min}$, which is a fraction of the total flux Φ according to Eq. 33:

33] $\Phi_x/\Phi = X_e/(X+X_{min})$

Combining Eqs. 32 and 33 yields:

34] $F = -\tfrac{1}{2}\Phi_x^2/\mu_o A$

Though the model of Eq. 34 is based on an empirical approximation, it reflects an approximate truth: that only the magnetic flux Φ_x passing through the armature results in force, while any additional flux associated with the winding flux linkage $n\Phi$ contributes to a stray inductance while contributing nothing to magnetic force.

X_e has been defined in terms of gap in a curve fit formula. In an operating solenoid, one computes X_e from the ratio of current to flux linkage, $I/n\Phi$. We start by recalling Eqs. 7 and 22:

7] $IL = n\Phi$

22] $X_e = \mu_o n^2 A/L$

Multiplying the numerator and denominator on the right of Eq. 22 by I, then substituting $n\Phi$ for IL in the denominator, and finally grouping terms yields:

35] $X_e = I\mu_o n^2 A/n\Phi = (I/n\Phi)(\mu_o n^2 A)$

Here we have the key equations for sensorless solenoid control. Eq. 4 shows that inductive voltage is the time derivative of flux linkage. Voltage going out through the leads of a solenoid can be controlled or measured. Current in the leads can be measured. These measurements take place on the controller circuit board, requiring no extra wiring to or sensors in the solenoid. Resistance in the solenoid winding and leads can be computed from voltage and current measurements made during periods when the current is steady, for example, when the solenoid is being held in a latched position. Under dynamic conditions, inductive voltage “Vi” is given by Vi = V – IR, applied voltage “V” minus measured current “I” multiplied by the self-calibrated resistance “R”. Starting from a known situation, for example at zero current and zero flux linkage for an open solenoid, the inductive voltage can be integrated over time to give a continuous measure of flux linkage, “$n\Phi$”. Eq. 35 shows that the ratio of measured current to the empirical flux linkage integral, “$I/n\Phi$”, is related by a constant scaling factor to “X_e”, the “effective” magnetic gap. Eq. 23 provides a simple empirical expression for relating “X_e” to the linearized position, “X”. This empirical expression has yielded excellent results in Magnesense work with linear solenoids and even with a rotary-motion valve rocker solenoid. Eq. 31 shows how to get from position and flux linkage to magnetic force. With the ability to compute magnetic force comes the ability to control force. A mass-spring-damper equation completes a good first pass at a complete electro-mechanical solenoid model. These are the basic relations needed to construct control algorithms. Other Magnesense papers and patents show how to follow through with the construction of practical solenoid control algorithms and hardware implementation.

Clearly, controlling a solenoid demands more than a good dynamic mathematical model. While various formulations may be equally “correct” in a mathematical sense, some formulations are far more robust than others for practical implementation of control. In practice, actually computing and making use of a numeric flux linkage integral for control purposes is a challenging task. Resistances vary with temperature. There is magnetic crosstalk between the two sides of a double-acting solenoid. Eddy currents and magnetic hysteresis, though “minor” in an efficient laminated solenoid, nevertheless call for first-order correction formulas, complicating the idealized formulas shown here. Obtaining good real-time sampled-data measurements in a noisy electromagnetic environment is challenging. Tracking flux linkage without excessive integration drift is a challenge, even for time periods as brief as a few milliseconds. With all these caveats and a few more, however, good results are attainable. Tracking flux linkage provides a robust starting point for dynamic solenoid control. The relationship of flux linkage and position to magnetic force is easier to work with, in controller practice, than the relationship of current and position to magnetic force, even though both pairs of variables carry the same information “in theory.” A flux linkage controller loop makes a more stable starting point for motion control than a current controller, because the relationship between flux linkage and force is better behaved.