Math. H110 Huge Generalized Inverses of Rank-Deficient Matrices Prof. W. Kahan

Normally we think of the solution $x = F^{-1}y$ of a linear system $Fx = y$ as a continuous function of the given data F and y . When no F^{-1} exists, any of a host of algebraic methods can be used to solve the system for x , at least approximately, although every solution's usefulness may be undermined by violent variability. This note assesses that violence quantitatively by relating it to the data's nearness to perturbed data of lower rank. Real number data is assumed.

Let F be a possibly rectangular matrix whose rank may be less than both of its dimensions for all we know. A *Generalized Inverse* G of F can be defined in many ways, all characterized thus: $x := Gy$ is a solution of the equation $Fx = y$ whenever this equation has a solution, even if the solution x is not determined uniquely by the equation. Consequently …

Lemma 0: The generalized inverses G of F are the solutions G of the equation $FGF = F$.

Proof: $y := Fx$ runs through F's range as x runs through its domain; then $x = Gy$ must be a solution of the equation $Fx = y$, whence $Fx = FGy = FGFx$ follows for all x , so $FGF = F$.

Lemma 1: Every matrix F , including $F = 0$, has at least one generalized inverse G .

This will be proved later. But first let us consider a widely used instance, the *Moore-Penrose Pseudo-Inverse* $F^\dagger$. It is the linear operator that solves a *Least*^{2}*-Squares* problem:

> Given F and y in F's target space (but perhaps not in F's range), find the vector x in F's domain that minimizes $\|Fx - y\|$ and, if the minimizing x is not yet unique, also minimizes $\|x\|$. Here $\|v\| := \sqrt{(v^Tv)}$ is the Euclidean length of real vector v .

Lemma 2: The Least^2-Squares problem's solution $x = F^\dagger y$ is obtainable from the *Singular-Value Decomposition* $F = QVP^T$, in which $Q^TQ = I$ and $P^TP = I$ and V is a nonnegative diagonal matrix (perhaps rectangular) exhibiting the singular values of F , by computing $F^\dagger := PV^\dagger Q^T$ where $V^\dagger$ is obtained from the diagonal V by transposing it and then replacing therein every nonzero diagonal element by its reciprocal. For example, $\begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}^\dagger = \begin{bmatrix} 1/3 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$.

Lemma 2 will be proved later. From its formula for pseudo-inverses will follow …

Lemma 3: $FF^\dagger F = F$, $F^\dagger FF^\dagger = F^\dagger$, $(F^\dagger F)^T = F^\dagger F$ and $(FF^\dagger)^T = FF^\dagger$; and these four equations characterize $F^\dagger$ because it is the unique matrix satisfying of all of them.

Lemma 3's first equation says that $F^\dagger$ is a generalized inverse of F , and the second says that F is a generalized inverse of $F^\dagger$; in fact, $(F^\dagger)^\dagger = F$. In general, generalized inverses are not reciprocal in this way; later we shall see how to construct a generalized inverse G of F that violates any chosen subset of the latter three equations of Lemma 3 unless $F = O$ or F has full rank (equal to F's least dimension). Unless F^{-1} exists (in which case $G = F^{-1}$), there are infinitely many generalized inverses G of F , almost all of them with gargantuan elements since $G+Z$ is also a generalized inverse for every Z satisfying $FZ = O$ or $ZF = O$.

Now, a gargantuan generalized inverse G is hazardous to approximate because subsequent computations of Gy may encounter massive cancellation leaving little more than error as a residue. And even when computation is performed exactly, or without roundoff, Gy can be changed utterly by tiny perturbations of y . Consequently, ample incentive exists to choose generalized inverses of near-minimal magnitudes; an instance is the Moore-Penrose Pseudo-Inverse, minimal because it solves the Least^2-Squares problem (Lemma 2). In fact, …

Lemma 4: Among generalized inverses G of F with minimal $\|G\| := \max_{y \neq o} \|Gy\|/\|y\|$ (the *Operator norm* based upon Euclidean vector norms $\|Gy\|$ and $\|y\|$), one turns out to be $F^\dagger$, and only it minimizes the *Root-Sum-Squares* norm $|G| := \sqrt{(\mathrm{Trace}(G^TG))} = \sqrt{(\sum_i \sum_j G_{ij}{}^2)}$.

The operator norm $\|G\|$ here turns out to be the biggest singular value of G , and $\|F^\dagger\|$ turns out to be the reciprocal of F's smallest nonzero singular value. $|F^\dagger|$ turns out to be the Root-Sum-Squares of reciprocals of all F's nonzero singular values. All this will become clear later.

Choosing a generalized inverse of near minimal magnitude cannot always avoid troubles with gargantuan magnitudes; they may be unavoidable if F is too near another matrix $F + \Delta F$ of lower rank because then every generalized inverse G of F must be huge. A quantitative statement of this cause-and-effect relationship involves matrix norms thus:

Theorem 5: Every generalized inverse G of F has

$$\|G\| \geq 1/\big(\ \min \|\Delta F\| \ \text{ such that } \ \mathrm{rank}(F - \Delta F) < \mathrm{rank}(F)\ \big) .$$

This will be proved later for *any* matrix norms $\|\ldots\|$ satisfying *Multiplicative Dominance* relations $\|Gy\| \leq \|G\| \cdot \|y\|$ and $\|Fx\| \leq \|F\| \cdot \|x\|$ required for *Compatibility* with vector norms; all plausible matrix norms are compatible that way. Theorem 5 says that the existence of at least one moderately sized generalized inverse of F implies that F is relatively well separated from all matrices $F - \Delta F$ of lower rank when separation is gauged by the chosen matrix norm.

Theorem 5 says nothing if the norms have been chosen to make F and G look good. For example, choose new coordinate bases in the domain and target spaces of F to represent it by a new matrix $\mathbb{F}$ consisting of an identity matrix bordered perhaps by zero matrices; the dimension of the identity matrix must equal the rank of F . Then $\mathbb{G} := \mathbb{F}^T$ is the representative in the new coordinates of a generalized inverse G of F in the old coordinates, and choosing *any* of the customary operator norms $\|\ldots\|$ in the new coordinates will induce corresponding operator norms $|\!|\ldots|\!|$ in the original coordinates such that $|\!|F|\!| := \|\mathbb{F}\| = 1$ and $|\!|G|\!| := \|\mathbb{G}\| = 1$. This is what is meant by norms "chosen to make F and G look good." Such a choice is not always praiseworthy.

Ideally, the chosen norm should serve the intended application. For instance, …

Ideally, all vector perturbations of roughly the same length,
gauged by the chosen norm,
should be roughly equally (in)consequential or (un)likely.

Worthwhile ideals are not easy to achieve; achieving this one, when possible, is a long story for another day. For now let us assume the norm is fixed in advance. For any given F we seek as small a generalized inverse G as can be computed at a tolerable cost. When the norm is the biggest-singular-value-norm used in Lemma 4, then a minimal generalized inverse G barely big enough to comply with Theorem 5 can be identified easily:

Corollary 6: The Moore-Penrose Pseudo-Inverse $F^\dagger$ is a generalized inverse of F satisfying

$$\|F^\dagger\| = 1/\big(\min \|\Delta F\| \text{ such that } \mathrm{rank}(F-\Delta F) < \mathrm{rank}(F)\big) ;$$

here $\|F^\dagger\|$ is the biggest singular value of $F^\dagger$, which is the reciprocal of F's smallest nonzero singular value which, in turn, is the distance from F to the nearest $F-\Delta F$ of lower rank.

Exercise: Show that $F^\dagger = \lim_{ß\to0+} (ßI+F^TF)^{-1}F^T = \lim_{ß\to0+} F(ßI+FF^T)^{-1}$.

For other operator norms $\|\ldots\|$ derived from non-Euclidean vector norms, a near minimal generalized inverse of $F \ne O$ can be difficult to construct unless F^{-1} exists, in which case …

Theorem 7: $\|F^{-1}\| = 1/\big(\min \|\Delta F\| \text{ such that } \det(F-\Delta F) = 0\big)$ for all operator norms $\|\ldots\|$.

Far harder than the proofs of the theorems, corollary and lemmas above is the proof of …

Theorem 8: For every operator norm $\|\ldots\|$, at least one generalized inverse $\overline{G}$ of F , besides satisfying Lemma 0 and Theorem 5, also satisfies

$$\|\overline{G}\| \le \sqrt{(\mathrm{rank}(F))}/\big(\min \|\Delta F\| \text{ such that } \mathrm{rank}(F-\Delta F) < \mathrm{rank}(F)\big) .$$

Theorem 8 shows that Theorem 5 is not excessively optimistic. I found a proof in 1973, but have not published it yet because I still hope some day to discover a much shorter proof.

• • • • • • • • • • • • • • • •

Here are proofs for the other assertions above. First is Lemma 1's assertion that every F has at least one generalized inverse G . One proof deduces that the equation FGF = F in Lemma 0 has a solution G by applying one of Fredholm's criteria: In general, the linear equation Ag = f must have at least one solution g if and only if $c^Tf = 0$ whenever $c^TA = o^T$. In our case, the linear operator A does to g what FGF does to G ; and c^Tf matches $\mathrm{Trace}(C^TF)$. In short, to apply Fredholm's criterion we must decide whether $\mathrm{Trace}(C^TF) = 0$ for every matrix C satisfying $\mathrm{Trace}(C^TFZF) = 0$ for all Z . Since the trace of a product is unchanged by cyclic permutation of its factors, the last equation implies that $\mathrm{Trace}(FC^TFZ) = 0$ for all Z , whence $FC^TF = O$, whence follows $(C^TF)^2 = O$, which implies that C^TF has no nonzero eigenvalue, whence follows that $\mathrm{Trace}(C^TF) = 0$ as required to establish that a solution G of FGF = F exists.

Another proof of G's existence follows changes of coordinate bases in the domain and target spaces of F to exhibit its linear operator's canonical form under *Equivalence*: $\overline{F} = \begin{bmatrix} I & O \\ O & O \end{bmatrix}$ wherein the dimension of the identity matrix I matches Rank(F) and the zero matrices O fill out the rest of the rows and columns; the rest of the rows or the rest of the columns or both can be absent if F has full rank equal to one of its dimensions. For these bases every generalized inverse $\overline{G} = \begin{bmatrix} I & R \\ S & Z \end{bmatrix}$ with arbitrary matrices R, S and Z so dimensioned (if not absent) that matrix products $\overline{F}\,\overline{G}$ and $\overline{G}\,\overline{F}$ both exist. Then $\overline{F}\,\overline{G}\,\overline{F} = \overline{F}$, and this equation persists in the form FGF = F after restoration of the original coordinate bases.

Exercise: Continuing from the last paragraph, prove that every generalized inverse G of F has Rank(G) ≥ Rank(F) with equality if and *only* if F is a generalized inverse of G too.

To prove Lemma 2 we use analogous coordinate changes, but this time to new orthonormal bases obtained from the orthogonal matrices P and Q of a singular value decomposition $F = Q\overline{F}P^T$ in which diagonal matrix $\overline{F} = \begin{bmatrix} V & O \\ O & O \end{bmatrix}$ has the same dimensions as matrix F whose singular values appear on $\overline{F}$'s diagonal; the square diagonal matrix V exhibits all nonzero singular values. We might as well assume $F = \overline{F}$ at the outset since these coordinate changes preserve Euclidean length in both domain and target spaces of F , leaving the Least^2-Squares problem unchanged. Now, with F diagonal, the Least^2-Squares problem's solution is almost obviously $x = F^\dagger y$ for $F^\dagger := \begin{bmatrix} V^{-1} & O \\ O & O \end{bmatrix}$ with F's dimensions transposed so that $F^\dagger F = \begin{bmatrix} I & O \\ O & O \end{bmatrix}$ is a square diagonal matrix, as is $FF^\dagger$. Also this diagonal $F^\dagger$ satisfies all four of Lemma 3's equations, which persist after restoration of the original coordinate bases. These equations determine $F^\dagger$ uniquely because they do so in the coordinate systems that diagonalize F , as is easily verified. In these coordinate systems, any $G = \begin{bmatrix} V^{-1} & R \\ S & Z \end{bmatrix}$ is a generalized inverse because it satisfies the first equation $FGF = F$ for any R, S and Z . Only $Z = -SV^{-1}R$ can satisfy the second equation $GFG = G$ too. Only $S = O$ can satisfy the first and third, $(GF)^T = GF$; only $R = O$ can satisfy the first and fourth, $(FG)^T = FG$. Consequently only $F^\dagger$ satisfies all four equations. This completes the proof of Lemma 3 and the second sentence after it.

When any generalized inverse G of F is represented as above in orthonormal coordinate bases that diagonalize F , the Root-Sum-Squares norm satisfies $|G|^2 = |V^{-1}|^2 + |R|^2 + |S|^2 + |Z|^2$, which exceeds $|F^\dagger|^2 = |V^{-1}|^2$ unless $G = F^\dagger$. This confirms the second assertion in Lemma 4. The first assertion's proof begins with the observation that clearing S and Z to zeros cannot increase $||G|| = \max_{y \neq o} ||Gy||/||y||$, nor can it be increased after that by clearing R ; therefore every generalized inverse G has its $||G|| \geq ||V^{-1}|| = ||F^\dagger|| = 1/$(F's least nonzero singular value) . This completes the proof of Lemma 4 and the paragraph after it.

Exercise: Fill in all unobvious details of the proofs above.

Exercise: Suppose that $F = LCR^T$ in which L, C and R all have the same rank and C is square and invertible. Show that $L^\dagger = (L^TL)^{-1}L^T$, $R^\dagger = (R^TR)^{-1}R^T$ and $F^\dagger = R^{\dagger T}C^{-1}L^\dagger$, so it is computable by rational arithmetic alone.

Current versions of Matlab compute an approximation `pinv(F)` to $F^\dagger$ from a singular value decomposition of F after all singular values tinier than a roundoff-related threshold have been reset to zeros. Otherwise reciprocals of these tiny singular values would introduce gargantuan numbers that would swamp everything else in $F^\dagger$ and render it useless numerically. A Matlab user can substitute his own threshold Ω for Matlab's by invoking `pinv(F, Ω)` . This is tantamount to perturbing F , changing it to $F-\Delta F$ with $||\Delta F|| < \Omega$ to get $||(F-\Delta F)^\dagger|| \leq 1/\Omega$, provided $\Omega > 0$ of course. If F has no singular values below that threshold, $\Delta F = O$; but otherwise $\text{Rank}(F-\Delta F) < \text{Rank}(F)$ and $||F^\dagger|| > 1/\Omega$. How should the threshold Ω be chosen?

If insights into data revealed by computed results are not to be confounded by an accident of the computational algorithm, then these results must be insensitive to ostensibly small changes in the threshold. This will be the case only if Ω falls into a relatively wide gap between F's small singular values and much tinier ones tiny enough to be discarded. Otherwise the choice of Ω becomes problematical.

Singular value decompositions reveal almost everything knowable about linear operators from one Euclidean space to another. What if the spaces are not both Euclidean? Vectors' lengths may be gauged by any of various norms $\|\dots\|$ each of which satisfies all the familiar laws

$$\|v\| > 0 \text{ except } \|o\| = 0 , \quad \|\mu\cdot v\| = |\mu|\cdot\|v\| , \quad \text{and} \quad \|u+v\| \le \|u\| + \|v\| ,$$

but violates the Euclidean norm's *Parallelogram Law* $\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2$; see our class notes on "How to Recognize a Quadratic Form". This violation by $\|\dots\|$ deprives the vector space of a wealth of *Isometries* (length-preserving linear transformations, the rotations and reflections represented by orthogonal matrices) possessed by Euclidean spaces. Here are two instances: The only isometries available for the biggest-magnitude norm and for the sum-of-magnitudes norm are the permutations and the sign reversals of column-vectors' elements. This is why non-Euclidean normed spaces (they are called "Banach spaces") have turned out much more difficult to analyse in the course of about a century of study.

The singular value decomposition has no useful counterpart for linear operators between spaces that are not both Euclidean, though the spaces' *Operator norm* $\|F\| := \max_{v\ne o} \|Fv\|/\|v\|$ does resemble the biggest singular value in some respects, and is easier to compute for the biggest-magnitude and sum-of-magnitudes vector norms provided both spaces use the same norm. All operator norms satisfy the product identity $\|uw^T\| = \|u\|\cdot\|w^T\|$ for rank-1 operators, as well as a multiplicative dominance relation $\|L\cdot F\| \le \|L\|\cdot\|F\|$ satisfied by all well-founded matrix norms.

Exercise: Prove these, recalling that the dual space's vector norm is an operator norm $\|w^T\| := \max_{v\ne o} |w^Tv|/\|v\|$. Symmetrically $\|v\| = \max_{w^T\ne o^T} |w^Tv|/\|w^T\|$; use this to prove all operator norms $\|F\| = \max_{w^T\ne o^T} \|w^TF\|/\|w^T\|$ too.

Exercise: Prove $\|Fx\| \le |F|\cdot\|x\|$ (compatibility) and $|FZ| \le |F|\cdot|Z|$ (multiplicative dominance) for $|\dots|$, though …

Some matrix norms are not operator norms; the root-sum-squares norm $|F|$ is an instance.
Here is why: Were there a pair of vector norms for which $|F|$ is the operator norm, it would satisfy the product identity $|uw^T| = \|u\|\cdot\|w^T\|$ for vectors' norm $\|\dots\|$ in the operators' target space and functionals' norm $\|\dots^T\|$ in the domain's dual space. Actually $|uw^T| = \sqrt{\mathrm{Trace}(wu^Tuw^T)} = \|u\|\cdot\|w^T\|$ for Euclidean norms in both spaces, and their operator norm is the biggest-singular-value, which is less that the root-sum-squares of singular values for all but rank-1 operators. In short, $|\dots|$ is generally too big to be an operator norm. In general, a matrix norm that is not an operator norm, but is *Compatible* (satisfies the multiplicative dominance relation) with the vector norms in the domain and target space, can be proved always at least as big as the operator norm for those two spaces.

Exercise: Prove that the Sum-of-all-magnitudes norm $|||F||| := \sum_i \sum_j |f_{ij}|$ cannot be an operator norm for 2-by-2 matrices $F = \{f_{ij}\}$, though it is compatible with the sum-of-magnitudes norm in its target space, and the biggest-element norm in its domain. Hint: try $F = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. (The operator norm for those two spaces is tedious to compute.)

Exercise: The Biggest-of-all-elements norm $\lceil G\rceil ::= \max_{ij} |g_{ij}|$ is an operator norm; for which vector norms and why? What is the least constant μ for which matrix norm $\mu\cdot\lceil G\rceil$ is compatible with the biggest-element norm in both G's domain and target spaces? Similarly, what … the sum-of-magnitudes norm in both … spaces?

Exercise: Explain why, if B is the matrix of a linear operator from a normed space to itself with a compatible $\|B\| < 1$, then I–B is nonsingular. However, if the square matrix B belongs to a linear operator beween two different normed spaces, then I–B can be singular despite that a compatible $\|B\| < 1 = \|I\|$; give an example.

Consider now two normed vector space V and W, and two matrix norms, one for matrices F that map V to W, and a second norm for matrices G that map W to V; and suppose both matrix norms are compatible with the spaces' vector norms: $\|Fv\| \le \|F\|\cdot\|v\|$ for all v in V, and $\|Gw\| \le \|G\|\cdot\|w\|$ for all w in W. (Were a matrix norm incompatible, multiplying it by a scalar constant big enough would make it compatible; do you see how?) For such norms, a lower bound for all generalized inverses of F comes from Theorem 5. Here is its proof:

So long as $\mathrm{Rank}(F) > \mathrm{Rank}(F-\Delta F) \ge \mathrm{Rank}((F-\Delta F)GF)$, the null-space of $(F-\Delta F)GF$ must be a subspace of F's domain with greater dimension than the null-space of F. Therefore vectors x must exist satisfying $(F-\Delta F)GFx = o \ne Fx$, whence follows $o \ne Fx = FGFx = \Delta F\cdot GFx$ and then $0 < \|Fx\| = \|\Delta F\cdot GFx\| \le \|\Delta F\|\cdot\|GFx\| \le \|\Delta F\|\cdot\|G\|\cdot\|Fx\|$ because of compatibility. Divide out $\|Fx\|$ and then minimize $\|\Delta F\|$ subject to $\mathrm{Rank}(F+\Delta F) < \mathrm{Rank}(F)$ to complete the proof of Theorem 5. It combines with Lemma 4 to yield Corollary 6.

Historically Corollary 6 is several decades older than Theorem 5, which seems three or four decades old. (See G.W. Stewart "On the Early History of the Singular Value Decomposition" in *SIAM Review* **35** (1993) pp. 551-566, for more chronology.) Both statements generalize Theorem 7, said to have been known to S. Banach in the 1920s, certainly known to M.G. Krein in the 1940s, and resurrected by numerical analyst N. Gastinel in the early 1960s.

For Theorem 7's proof we assume that F is a conventionally invertible square matrix, and we seek the singular matrix $F-\Delta F$ nearest F in the sense that operator norm $\|\Delta F\|$ is minimized. Theorem 5 says the minimum can't be smaller than $1/\|F^{-1}\|$, so constructing ΔF to achieve equality will prove the desired result.

(Note that the notation used for norms here is still, as usual, "overloaded" because $\|\Delta F\|$ and $\|F^{-1}\|$ can be gauged by different operator norms when the spaces between which F and F^{-1} operate (in opposite directions) are gauged by different vector norms. None the less, you should be able to see why $\|F\|\cdot\|F^{-1}\| \ge \|I\| = 1$; try it!.)

We will devise a minimizing $\Delta F := vw^T$ of rank 1 as follows: Since $\|F^{-1}\| = \max_{\|x\|=1} \|F^{-1}x\|$ let $x = v$ be a maximizing vector; then $\|F^{-1}v\| = \|F^{-1}\|$ and $\|v\| = 1$. We also know that $\|F^{-1}v\| = \max_{\|y^T\|=1} y^T(F^{-1}v)$, and can now choose a maximizing functional $y^T = u^T$; then $\|F^{-1}v\| = u^T(F^{-1}v)$ and $\|u^T\| = 1$. Finally set $w^T := u^T/\|F^{-1}\|$ to get $w^TF^{-1}v = 1$. Now $\Delta F := vw^T$ has $\|\Delta F\| = \|v\|\cdot\|w^T\| = 1/\|F^{-1}\|$ and $(F-\Delta F)F^{-1}v = v - vw^TF^{-1}v = o$, which implies that $F-\Delta F$ is a singular matrix nearest F. Theorem 7's proof ends.

Alas, my proof of Theorem 8 is much too long to reproduce here.

• • • • • • • • • • • • • • • •

Perturbations

How does its generalized inverse change with F ? Put this way, the question begs a crucial question: Is “its generalized inverse” determined uniquely by F ? It is in important special cases. For instance, an easily verified identity satisfied by any invertible square matrix F is

$$(F+\delta F)^{-1} - F^{-1} = -(F+\delta F)^{-1}\cdot\delta F\cdot F^{-1} = -F^{-1}\cdot\delta F\cdot(F+\delta F)^{-1}$$

provided $(F+\delta F)^{-1} = (I + F^{-1}\cdot\delta F)^{-1}F^{-1}$ exists too, as it must when $\|F^{-1}\cdot\delta F\| < 1$ for a suitable (see the previous exercise) operator norm, as is surely the case when $\|F^{-1}\|\cdot\|\delta F\| < 1$. (Why?) Then $\|(F+\delta F)^{-1} - F^{-1}\|/\|F^{-1}\| \le 1/(1/\|F^{-1}\delta F\| - 1)$, with equality for an apt δF of rank 1 .

Exercise: Confirm the last sentence. Then exhibit a $dF/d\tau$ for which $\|d(F^{-1})/d\tau\| = \|F^{-1}\|^2\cdot\|dF/d\tau\| \ne 0$.

Thus F^{-1} and its derivative $dF^{-1}/d\tau$ become huge together only as an invertible F approaches some singular matrix. To behave analogously when F is not invertible, its generalized inverse must be determined uniquely by some conditions besides the one in Lemma 0. Non-metric (no norms) conditions that sometimes determine a generalized inverse G of F uniquely do exist.

For instance, when F is square the three equations $FGF = F$, $GFG = G$ and $FG = GF$ always have at most one solution G , but sometimes none; and when a solution G exists it can vary arbitrarily violently as F changes even though Rank(F) does not change. I have heard this generalized inverse G called “Drazin’s Semi-Inverse” when it exists, which it does if and only if $Fz = o$ whenever $F^kz = o$ for any integer $k > 0$. This means G exists if and only if F’s *Jordan Normal Form* has no nonzero Jordan block with diagonal all zero. To see why, suppose $F^kz = o$ for some $k > 0$, and that a solution G exists. Then $Gz = (GF)^kGz = G^{k+1}F^kz = o$, whence $Fz = F^2Gz = o$ too. This restricts Jordan’s Normal Form of F to have no eigenvalue 0 with a Jordan block bigger than 1-by-1 . Conversely, suppose the Jordan blocks of F are constrained that way. The characteristic polynomial of F is then $\det(ßI - F) = \sum_{0\le j\le m} \mu_j ß^{j+k}$ with $\mu_m = 1$, $\mu_0 \ne 0$ and some $k \ge 0$, and the Cayley-Hamilton Theorem ensures that this polynomial vanishes when F is substituted for ß . The constraint implies $\sum_{0\le j\le m} \mu_j F^{j+K} = O$ with $K := \min\{k, 1\} = 0$ or 1 . Now $G := (\sum_{1\le j\le m} \mu_j F^{j-1}/\mu_0)^2F$ turns out to be the solution of the three equations in question; can you confirm this? There is no other solution because, if Z also satisfies the three equations $FZF = F$, $ZFZ = Z$ and $FZ = ZF$, then $Z = (ZF)^2Z = Z^3F^2 = Z^3F^4G^2 = (ZF)^3FG^2 = FG^2 = G$; it is unique. Note that this solution G (when it exists) is a rational function of F .

Exercise: Given a rank-1 square matrix $F = uv^T$ with Euclidean $\|u\| = \|v^T\| = 1$ and $v^Tu = \cos(\theta) \ne 0$, show that $F^\dagger = F^T$ but Drazin’s semi-inverse $G = F/\cos^2(\theta)$. Evidently it can be enormously bigger than $F^\dagger$.

Among uniquely defined generalized (not ordinary) inverses, the Moore-Penrose Pseudo-Inverse $F^\dagger$ is the most commonly used. How does it change when F is perturbed? Because $F^\dagger$ is a rational function of F , formulas generalizing the identity near the top of this page must exist presenting the change in a way that allows a limit process to express the derivative $dF^\dagger/d\tau$ (when it exists) in terms of $dF/d\tau$. Here is such a formula (with E in place of $F+\delta F$):

Lemma 9: $E^\dagger - F^\dagger = -F^\dagger(E-F)E^\dagger + (I - F^\dagger F)(E-F)^TE^{\dagger T}E^\dagger + F^\dagger F^{\dagger T}(E-F)^T(I - EE^\dagger)$.

Proof: The identity’s right-hand side expands into ten terms. Two of them condense, as does $F^\dagger F^{\dagger T}F^T$ to $F^\dagger$, and persist. Four of them condense, as does $F^\dagger FF^TE^{\dagger T}E^\dagger$ to $F^TE^{\dagger T}E^\dagger$, and cancel the remaining four terms, thus confirming the identity. I presented it in 1971 at an IFIP Congress in Ljubljiana, but it had already been discovered in Lund by P-Å. Wedin for his 1969 thesis, most of which he published in 1973 in *BIT* **13**.

Lemma 9 leads to the following overestimate of the biggest-singular-value norm of $E^\dagger – F^\dagger$:

Theorem 10: If $E \neq F$ then

$$\|E^\dagger - F^\dagger\|/\|E–F\| \le \sqrt{(\ \|E^\dagger\|^4 + \|E^\dagger\|^2 \cdot \|F^\dagger\|^2 + \|F^\dagger\|^4\ )} \ \le\ \sqrt{3} \cdot \max\{\|E^\dagger\|, \|F^\dagger\|\}^2\ .$$

Proof: Lemma 9's identity exhibits $E^\dagger - F^\dagger = R + L - S$ wherein $S := F^\dagger(E–F)E^\dagger$, $R := \Phi(E–F)^T E^{\dagger T} E^\dagger$ and $L := F^\dagger F^{\dagger T}(E–F)^T \Psi$ in which the orthogonal projector $\Phi := I - F^\dagger F = \Phi^T = \Phi^2$ satisfies $F\Phi = O = \Phi F^\dagger$ and $\|\Phi\| = 1$ except when $\Phi = O$; similarly for $\Psi := I - EE^\dagger$. Now we can estimate $\|E^\dagger - F^\dagger\|$ by using the easily verified formula $\|Z^T\|^2 = \|Z\|^2 = \|Z^T Z\| =$ (the biggest eigenvalue of $Z^T Z$) . Since $R^T S = R^T L = O$ we find $(E^\dagger - F^\dagger)^T(E^\dagger - F^\dagger) = R^T R + (L–S)^T(L–S)$, whence $\|E^\dagger - F^\dagger\|^2 \le \|R\|^2 + \|L–S\|^2$. And since $LS^T = O$ we find $(L–S)(L–S)^T = LL^T + SS^T$ whence $\|L–S\|^2 \le \|L\|^2 + \|S\|^2$. Because $\|R\| \le 1 \cdot \|E–F\| \cdot \|E^\dagger\|^2$, $\|L\| \le \|F^\dagger\|^2 \cdot \|E–F\| \cdot 1$ and $\|S\| \le \|F^\dagger\| \cdot \|E–F\| \cdot \|E^\dagger\|$, putting it all together yields $\|E^\dagger - F^\dagger\|/\|E–F\| \le \sqrt{(\ \|E^\dagger\|^4 + \|E^\dagger\|^2 \cdot \|F^\dagger\|^2 + \|F^\dagger\|^4\ )}$ as claimed. P-Å. Wedin's more penetrating proof got a more complicated estimate with $(1+\sqrt{5})/2$ in place of $\sqrt{3}$.

$E^\dagger := (F+\delta F)^\dagger$ can change violently for a very tiny δF if $\mathrm{Rank}(F+\delta F) > \mathrm{Rank}(F)$ since then Corollary 6 implies $\|(F+\delta F)^\dagger\| \ge 1/\|\delta F\|$, rendering Theorem 10's overestimate gargantuan and useless. On the other hand, when $\mathrm{Rank}(F+\delta F) = \mathrm{Rank}(F)$ and $\|\delta F\| < 1/\|F^\dagger\|$, so $\|\delta F\|$ is too tiny for any perturbation of its size to drop the rank of $F+\delta F$, then …

Lemma 11: If positive, $\|F^\dagger\|/(1 - \|F^\dagger\| \cdot \|\delta F\|) \ge \|(F+\delta F)^\dagger\|$ provided $\mathrm{Rank}(F+\delta F) = \mathrm{Rank}(F)$ too.

Proof: Let $F–\Delta F$ be the matrix of rank less than $\mathrm{Rank}(F+\delta F)$ nearest $F+\delta F$ when gauged by the Biggest-singular-value norm. Then Corollary 6 implies both $\|\delta F+\Delta F\| = \|F+\delta F - F+\Delta F\| = 1/\|(F+\delta F)^\dagger\|$ and $1/\|F^\dagger\| \le \|\Delta F\| = \|\delta F+\Delta F - \delta F\| \le \|\delta F+\Delta F\| + \|\delta F\| = 1/\|(F+\delta F)^\dagger\| + \|\delta F\|$, which turns into the lemma's inequality. Inequalities more general than this, because they allow δF to be somewhat bigger, and sharper than this and Theorem 10's inequalities can be found in Wedin (1973) *BIT* (*Nordisk Tidskrift for Informationsbehandling*) **13** pp. 217-232, and in Stewart (1977) *SIAM Review* **19** pp. 634-662 which surveys the subject in depth. They go far deeper than necessary for this course.

Lemma 11 and Theorem 10 imply that $\|(F+\delta F)^\dagger - F^\dagger\| \le \sqrt{3}\|\delta F\| \cdot \|F^\dagger\|^2/(1 - \|F^\dagger\| \cdot \|\delta F\|)^2$ so long as $\mathrm{Rank}(F+\delta F) = \mathrm{Rank}(F)$ and $\|\delta F\| < 1/\|F^\dagger\|$. Together with Lemma 9, these yield …

Theorem 12: Provided $\mathrm{Rank}(F)$ does not change as F varies, $F^\dagger$ is a continuously differentiable rational function of F with

$$dF^\dagger/d\tau = -F^\dagger(dF/d\tau)F^\dagger + (I - F^\dagger F)(dF/d\tau)^T F^{\dagger T} F^\dagger + F^\dagger F^{\dagger T}(dF/d\tau)^T(I - FF^\dagger)\ , \quad \text{and}$$

$$\|dF^\dagger/d\tau\| \le \sqrt{3} \cdot \|dF/d\tau\| \cdot \|F^\dagger\|^2\ ,$$

and, since $\|F^\dagger\| = 1/$(distance from F to the nearest matrix of lower rank) , neither $\|F^\dagger\|$ nor $\|dF^\dagger/d\tau\|/\|dF/d\tau\|$ can become gargantuan unless F approaches a matrix of lower rank.

This illustrates an important general theme among algebraic problems. Each such problem may be embedded in a family of more general problems: a linear system of equations, or a system of polynomial equations. If such an embedding loses sight of a significant combinatorial attribute of the given problem, like the rank of a matrix, or like the multiplicity of an eigenvalue or a polynomial's zero, the admission of even infinitesimal perturbations that upset that combinatorial attribute can turn a tame problem into a knotty (if not naughty) one. The threshold Ω in Matlab's `pinv` provides a way to restore a matrix's rank after it was upset by rounding errors; if successful, this compensation improves the computed result enormously. If compensation fails, serious rethinking is in order.