THE MANY DISGUISES OF THE SCHUR COMPLEMENT

ANTOINE LEVITT

Contents

The Schur complement is a simple technique in linear algebra which reduces the number of degrees of freedom of a problem by explicitly solving a part of it. It is at the heart of a very large number of seemingly unrelated techniques in various fields of mathematics and applications. This short note aims at presenting some of these many techniques in a unified way. Nothing here is new, and much more detail can be found in specialized sources. This note is by nature incomplete and biased towards my own interests. Please contact me if you have additional examples!

For simplicity most results are stated in finite-dimensional spaces and in a somewhat careless fashion, but the analysis can be conducted rigorously. It also extends to infinite-dimensional spaces with the appropriate hypotheses (and in fact the concept of Schur complement is often used to reduce a problem from an infinite-dimensional setting to a finite-dimensional one).

1. The Schur complement

The basic framework of Schur complements is that of an operator A acting on a space $X = X_1 \oplus X_2$ (that is, a partitioning of the degrees of freedom in two disjoint subsets 1 and 2). Often this decomposition is orthogonal but it needs not be. We can write the operator A in block form as

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}. \tag{1}$$

Assume now that we want to solve the linear system $Ax = b$. Then we can write it as

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 &= b_1 \\ A_{21}x_1 + A_{22}x_2 &= b_2 \end{aligned}$$

Assuming that A_{22} is invertible, we can solve x_2 as a function of x_1:

$$x_2 = A_{22}^{-1}(b_2 - A_{21}x_1)$$

and replace in the first equation to obtain

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})x_1 = b_1 - A_{12}A_{22}^{-1}b_2 \tag{2}$$

The operator

$$S = A_{11} - A_{12}A_{22}^{-1}A_{21} \tag{3}$$

is called the Schur complement.

It follows from the steps above that, assuming that A_{22} is invertible, A is invertible if and only if S is. This can be strengthened in various ways. For instance, since $Ax = b$ is the optimality condition of the functional $\frac{1}{2}x^T Ax - b^T x$, it follows that if A_{22} is symmetric definite positive, A

is symmetric definite positive if and only if S is; in this case, $x_2 = A_{22}^{-1}(b_2 - A_{21}x_1)$ is the result of minimizing the functional with respect to x_2, at fixed x_1. Another useful generalization is the Haynsworth inertia formula, which relates the number of positive and negative eigenvalues of A to those of S and A_{22}.

2. Linear algebra: block Gaussian elimination

The Schur complement can be viewed as a block generalization of the venerable Gaussian elimination. Recall that Gaussian elimination in the 2×2 case starts from the equations

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 &= b_1 \\ A_{21}x_1 + A_{22}x_2 &= b_2 \end{aligned}$$

where A_{ij} are scalars, and modifies the first equation (usually this process is done in reverse; we reverse it here for consistency) by

$$L_1 \leftarrow L_1 - A_{12}A_{22}^{-1}L_2$$

to remove x_2 from the first equation, simply leaving (2).

3. Linear algebra: the Sherman-Morrison formula

This is not a straightforward Schur complement formula, but has a very similar flavor. Assume that A is invertible, and let u be a vector. What is the inverse of the rank-1 update $A + uu^T$? Write the linear system as

$$Ax + uu^Tx = b$$

Following the idea of the Schur complement, we try to solve parts of the equation as a function of other parts of the solution itself:

$$x = A^{-1}(b - uu^Tx)$$

The solution is known up to the scalar $s = u^Tx$ (which plays the same role as x_1 in the Schur complement). To determine this scalar, we take the inner product with u and obtain

$$\begin{aligned} s &= u^TA^{-1}b - su^TA^{-1}u \\ s &= \frac{u^TA^{-1}b}{1 + u^TA^{-1}u} \end{aligned}$$

and therefore

$$(A + uu^T)^{-1} = A^{-1} - \frac{A^{-1}uu^TA^{-1}}{1 + u^TA^{-1}u}$$

This can be extended to rank-n (where it is known as the Woodbury formula) and non-symmetric updates by the same method.

4. Nonlinear analysis: the Lyapunov-Schmidt reduction

The Lyapunov-Schmidt reduction locally reduces the study of a branch of solutions to a nonlinear equation to a nonlinear system of dimension as small as possible. It is a nonlinear form of Schur complementation.

Consider the equation

$$f(x, \lambda) = 0$$

where $f : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ is smooth. Assume that $f(0,0) = 0$; we are interested in the behavior of the solutions of this equation close to $(0,0)$. From the implicit function theorem, when $J = \partial_x f$ is invertible the equation $f(x,\lambda) = 0$ has a locally unique solution for λ small. What happens when this is not true?

Partition the input space as $X_1 \oplus X_2$ where $X_1 = \mathrm{Ker}(J)$, and the output space as $F_1 \oplus F_2$ where $F_2 = \mathrm{Ran}J$. Then

$$J = \begin{pmatrix} 0 & 0 \\ 0 & J_{X_2F_2} \end{pmatrix}.$$

The operator $J_{X_2F_2}$ is injective and surjective from X_2 to X_2, and therefore invertible. Write $f(x,\lambda) = 0$ as

$$\begin{aligned} P_{F_1} f(x_1 + x_2, \lambda) &= 0 \\ P_{F_2} f(x_1 + x_2, \lambda) &= 0 \end{aligned}$$

From the implicit function theorem, the second equation can be locally solved for x_2 as a function of x_1: $x_2(x_1)$. Plugging back into the first equation, we get

$$P_{F_1} f(x_1 + x_2(x_1), \lambda) = 0$$

where P_{F_i} are the projections on F_i with respect to the decomposition $F = F_1 \oplus F_2$. This is now a purely nonlinear equation (no linear terms) for x_1 only. For instance, when $X_1 = \ker J$ is of dimension 1, this is a scalar equation, known as the bifurcation equation. The study of this scalar equation (usually through the higher derivatives) yields the local behavior of the solutions (saddle-node, pitchfork, etc.)

5. Spectral theory: the Feshbach-Schur method

This method can be seen as a special case of the Lyapunov-Schmidt reduction applied to the spectral problem $A_\mu x = \lambda x$, where μ is a parameter. Assuming that $A_{\mu,22} - \lambda$ is invertible, λ is an eigenvalue of A_μ if and only if the "Feshbach-Schur map"

$$S_\mu(\lambda) = A_{\mu,11} - \lambda - A_{\mu,12}(A_{\mu,22} - \lambda)^{-1} A_{\mu,21}$$

is singular. Assume for instance that A is hermitian and that x_0 is a simple eigenvector of A_0. Then, choosing $X_1 = \mathrm{Span}(x_0)$, $S_\mu(\lambda) = 0$ is a scalar equation which can be studied using regular finite-dimensional perturbation theory.

6. Optimization: saddle point problems

Consider the following optimization problem:

$$\begin{aligned} &\min f(x) \\ &\text{s.t. } g(x) = 0 \end{aligned}$$

Then if the jacobian of g at a constrained minimum x_* is full-rank, there is a set of Lagrange multipliers λ_* such that (x_*, λ_*) is a critical point of the Lagrangian

$$L(x,\lambda) = f(x) - \lambda^T g(x)$$

The Hessian is

$$\nabla^2 L(x,\lambda) = \begin{pmatrix} \nabla^2 f - \lambda^T \nabla^2 g & -\nabla g \\ -\nabla g^T & 0 \end{pmatrix}$$

Assume now that $\nabla^2 f - \lambda^T \nabla^2 g$ is positive definite at (x_*, λ_*). It follows from the implicit function theorem that the map

$$h(\lambda) = \min_x L(x,\lambda)$$

is defined for λ close to λ_*, where in this formula the search for x is reduced to a neighborhood of x_*. By the Hellmann-Feynmann/envelope theorem, $\nabla h(\lambda_*) = g(x_*) = 0$, and

$$\nabla^2 h(\lambda_*) = -\nabla g^T (\nabla^2 f - \lambda^T \nabla^2 g)^{-1} \nabla g$$

which is nothing but the Schur complement of $\nabla^2 L(x,\lambda)$. This is a negative definite matrix, which shows that λ_* is a local maximum for $h(\lambda)$ (a saddle point structure).

7. Electrical engineering: equivalent circuits

The problem of solving a circuit composed of resistors, inductors, capacitances and current sources in the harmonic regime reduces to solving an equation

$$A(\omega)x(\omega) = b(\omega)$$

where x are the voltages at the nodes of the circuit, b is the current sources, A is the conductance matrix and ω is the frequency. Assume now that a part of this circuit is a subcircuit, which has a number of nodes ("ports") by which it communicates to the rest of the circuit, and a number of internal nodes that do not communicate with the rest. Let X_2 be the space corresponding to these internal nodes. Then one can solve for the internal nodes as a function of the values of the voltages at the ports. The resulting equation for the non-internal nodes x_1 is

$$(A_{11}(\omega) - A_{12}(\omega)A_{22}(\omega)^{-1}A_{21}(\omega))S(\omega)x_1(\omega) = b_1(\omega) - A_{12}(\omega)A_{22}(\omega)^{-1}b_2(\omega)$$

Therefore, the subcircuit can equivalently be thought of as an equivalent circuit that links the ports with a conductance matrix $A_{12}(\omega)A_{22}(\omega)^{-1}A_{21}(\omega)$ and a current source $A_{12}(\omega)A_{22}(\omega)^{-1}b_2(\omega)$. This is known as the Kron reduction.

8. Quantum mechanics: the self-energy

This is an application of the Feshbach-Schur technique. Let H be the Hamiltonian of a quantum system 1 in interaction with an "environment" 2. Then the Schrödinger equation $H\psi = \lambda\psi$ can be rewritten as

$$(H_{11} + \Sigma(\lambda))\psi_1 = \lambda\psi_1$$

where

$$\Sigma(\lambda) = -H_{12}(H_{22} - \lambda)^{-1}H_{21}$$

is called the "self-energy". The physical interpretation is that, through coupling with the system 2, the system 1 acquires an effective additional energy. The total Hamiltonian $H_{11} + \Sigma(\lambda)$ describes a "dressed" system 1, which includes the reaction of the environment. In particular, when H_{22} has continuous spectrum, $\lim_{\eta \to 0^+} \sum(\omega + i\eta)$ can have a non-self-adjoint part, which describes (in a specific regime) irreversible transfer from the system 1 to the environment (Fermi golden rule).

9. Partial differential equations: the Stokes equation

Consider a bounded domain Ω, and the stationary adimensional Stokes equation for the velocity u and pressure p

$$\nabla \cdot u = 0$$
$$\nabla p - \Delta u = 0$$

with appropriate boundary conditions. In matrix form, this is

$$A = \begin{pmatrix} 0 & \nabla^T \\ \nabla & -\Delta \end{pmatrix}$$

If $-\Delta$ is invertible (which is the case for example with Dirichlet boundary conditions), the Schur complement reduces the study of the original problem to that of the operator $\nabla^T(-\Delta)^{-1}\nabla$ on pressure variables; its invertibility is sometimes called the inf-sup condition, and plays an important role in the study of finite element discretizations.

10. Partial differential equations: the Dirichlet-to-Neumann map

Consider the problem of solving an elliptic partial differential equation

$$Lu = f$$

on the whole space. Assume f is zero outside a domain Ω of interest. Two types of boundary data can be prescribed on $\partial\Omega$: Dirichlet data $u|_{\partial\Omega}$, and Neumann data $(\partial u/\partial n)|_{\partial\Omega}$, where n is the outgoing normal to Ω. Call 1 the degrees of freedom inside of Ω, and 2 those outside. If the Dirichlet data of u is known on Ω, then the equation $Lu_2 = 0, u_2|_{\partial\Omega} = g$ can be solved for all functions g on $\partial\Omega$. Let S be the Dirichlet-to-Neumann map, the operator that to g maps $(\partial u_2/\partial n)_{\partial\Omega}$. Then u_1 satisfies the equation

$$Lu_1 = f, \quad (\partial u_1/\partial n)|_{\partial\Omega} = Su_1$$

which is posed on the inside of Ω, at the price of a modified (possibly nonlocal) boundary condition.

The connection to Schur complements is even more manifest at the discrete level. The degrees of freedom can then be partitioned into those inside Ω and those outside. The Schur complement $L_{11} - L_{12}L_{22}^{-1}L_{21}$ is a nonzero modification of L_{11} only on the boundary, which can be interpreted as a modified boundary condition. This is the basis of some domain decomposition schemes.

Similar schemes can be used to reduce the solution of a volumic problem to that of a surface problem (e.g. boundary element method). In all cases it is important to be able to solve the exterior problem more or less explicitly, for instance through its Green function.