12-10-2010

Notes on the Schur Complement

Jean H. Gallier
University of Pennsylvania, jean@cis.upenn.edu

Gallier, J., Notes on the Schur Complement

The Schur Complement and Symmetric Positive Semidefinite (and Definite) Matrices

Jean Gallier

December 10, 2010

1 Schur Complements

In this note, we provide some details and proofs of some results from Appendix A.5 (especially Section A.5.5) of *Convex Optimization* by Boyd and Vandenberghe [1].

Let M be an $n \times n$ matrix written a as 2×2 block matrix

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where A is a $p \times p$ matrix and D is a $q \times q$ matrix, with $n = p + q$ (so, B is a $p \times q$ matrix and C is a $q \times p$ matrix). We can try to solve the linear system

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} c \\ d \end{pmatrix},$$

that is

$$\begin{aligned} Ax + By &= c \\ Cx + Dy &= d, \end{aligned}$$

by mimicking Gaussian elimination, that is, assuming that D is invertible, we first solve for y getting

$$y = D^{-1}(d - Cx)$$

and after substituting this expression for y in the first equation, we get

$$Ax + B(D^{-1}(d - Cx)) = c,$$

that is,

$$(A - BD^{-1}C)x = c - BD^{-1}d.$$

If the matrix $A - BD^{-1}C$ is invertible, then we obtain the solution to our system

$$\begin{aligned} x &= (A - BD^{-1}C)^{-1}(c - BD^{-1}d) \\ y &= D^{-1}(d - C(A - BD^{-1}C)^{-1}(c - BD^{-1}d)). \end{aligned}$$

The matrix, $A - BD^{-1}C$, is called the *Schur Complement* of D in M. If A is invertible, then by eliminating x first using the first equation we find that the Schur complement of A in M is $D - CA^{-1}B$ (this corresponds to the Schur complement defined in Boyd and Vandenberghe [1] when $C = B^{\top}$).

The above equations written as

$$\begin{aligned} x &= (A - BD^{-1}C)^{-1}c - (A - BD^{-1}C)^{-1}BD^{-1}d \\ y &= -D^{-1}C(A - BD^{-1}C)^{-1}c + (D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1})d \end{aligned}$$

yield a formula for the inverse of M in terms of the Schur complement of D in M, namely

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{pmatrix}.$$

A moment of reflexion reveals that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & 0 \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} \end{pmatrix} \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix},$$

and then

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} I & 0 \\ -D^{-1}C & I \end{pmatrix} \begin{pmatrix} (A - BD^{-1}C)^{-1} & 0 \\ 0 & D^{-1} \end{pmatrix} \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix}.$$

It follows immediately that

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}C & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C & I \end{pmatrix}.$$

The above expression can be checked directly and has the advantage of only requiring the invertibility of D.

Remark: If A is invertible, then we can use the Schur complement, $D - CA^{-1}B$, of A to obtain the following factorization of M:

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} I & 0 \\ CA^{-1} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & D - CA^{-1}B \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & I \end{pmatrix}.$$

If $D - CA^{-1}B$ is invertible, we can invert all three matrices above and we get another formula for the inverse of M in terms of $(D - CA^{-1}B)$, namely,

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}.$$

If A, D and both Schur complements $A - BD^{-1}C$ and $D - CA^{-1}B$ are all invertible, by comparing the two expressions for M^{-1}, we get the (non-obvious) formula

$$(A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}.$$

Using this formula, we obtain another expression for the inverse of M involving the Schur complements of A and D (see Horn and Johnson [5]):

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} (A - BD^{-1}C)^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{pmatrix}.$$

If we set $D = I$ and change B to $-B$ we get

$$(A + BC)^{-1} = A^{-1} - A^{-1}B(I - CA^{-1}B)^{-1}CA^{-1},$$

a formula known as the *matrix inversion lemma* (see Boyd and Vandenberghe [1], Appendix C.4, especially C.4.3).

2 A Characterization of Symmetric Positive Definite Matrices Using Schur Complements

Now, if we assume that M is symmetric, so that A, D are symmetric and $C = B^\top$, then we see that M is expressed as

$$M = \begin{pmatrix} A & B \\ B^\top & D \end{pmatrix} = \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BD^{-1}B^\top & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix}^\top,$$

which shows that M is similar to a block-diagonal matrix (obviously, the Schur complement, $A - BD^{-1}B^\top$, is symmetric). As a consequence, we have the following version of "Schur's trick" to check whether $M \succ 0$ for a symmetric matrix, M, where we use the usual notation, $M \succ 0$ to say that M is positive definite and the notation $M \succeq 0$ to say that M is positive semidefinite.

Proposition 2.1 *For any symmetric matrix, M, of the form*

$$M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix},$$

if C is invertible then the following properties hold:

(1) $M \succ 0$ iff $C \succ 0$ and $A - BC^{-1}B^\top \succ 0$.

(2) If $C \succ 0$, then $M \succeq 0$ iff $A - BC^{-1}B^\top \succeq 0$.

Proof. (1) Observe that

$$\begin{pmatrix} I & BD^{-1} \\ 0 & I \end{pmatrix}^{-1} = \begin{pmatrix} I & -BD^{-1} \\ 0 & I \end{pmatrix}$$

and we know that for any symmetric matrix, T, and any invertible matrix, N, the matrix T is positive definite ($T \succ 0$) iff $NTN^\top$ (which is obviously symmetric) is positive definite ($NTN^\top \succ 0$). But, a block diagonal matrix is positive definite iff each diagonal block is positive definite, which concludes the proof.

(2) This is because for any symmetric matrix, T, and any invertible matrix, N, we have $T \succeq 0$ iff $NTN^\top \succeq 0$. □

Another version of Proposition 2.1 using the Schur complement of A instead of the Schur complement of C also holds. The proof uses the factorization of M using the Schur complement of A (see Section 1).

Proposition 2.2 *For any symmetric matrix, M, of the form*

$$M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix},$$

if A is invertible then the following properties hold:

(1) $M \succ 0$ iff $A \succ 0$ and $C - B^\top A^{-1} B \succ 0$.

(2) If $A \succ 0$, then $M \succeq 0$ iff $C - B^\top A^{-1} B \succeq 0$.

When C is singular (or A is singular), it is still possible to characterize when a symmetric matrix, M, as above is positive semidefinite but this requires using a version of the Schur complement involving the pseudo-inverse of C, namely $A - BC^\dagger B^\top$ (or the Schur complement, $C - B^\top A^\dagger B$, of A). But first, we need to figure out when a quadratic function of the form $\frac{1}{2}x^\top Px + x^\top b$ has a minimum and what this optimum value is, where P is a symmetric matrix. This corresponds to the (generally nonconvex) quadratic optimization problem

$$\text{minimize} \quad f(x) = \frac{1}{2}x^\top Px + x^\top b,$$

which has no solution unless P and b satisfy certain conditions.

3 Pseudo-Inverses

We will need pseudo-inverses so let's review this notion quickly as well as the notion of SVD which provides a convenient way to compute pseudo-inverses. We only consider the case of square matrices since this is all we need. For comprehensive treatments of SVD and pseudo-inverses see Gallier [3] (Chapters 12, 13), Strang [7], Demmel [2], Trefethen and Bau [8], Golub and Van Loan [4] and Horn and Johnson [5, 6].

Recall that every square $n \times n$ matrix, M, has a *singular value decomposition*, for short, *SVD*, namely, we can write

$$M = U\Sigma V^\top,$$

where U and V are orthogonal matrices and Σ is a diagonal matrix of the form

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0),$$

where $\sigma_1 \geq \cdots \geq \sigma_r > 0$ and r is the rank of M. The σ_i's are called the *singular values* of M and they are the positive square roots of the nonzero eigenvalues of $MM^\top$ and $M^\top M$. Furthermore, the columns of V are eigenvectors of $M^\top M$ and the columns of U are eigenvectors of $MM^\top$. Observe that U and V are not unique.

If $M = U\Sigma V^\top$ is some SVD of M, we define the *pseudo-inverse*, $M^\dagger$, of M by

$$M^\dagger = V\Sigma^\dagger U^\top,$$

where

$$\Sigma^\dagger = \mathrm{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0).$$

Clearly, when M has rank $r = n$, that is, when M is invertible, $M^\dagger = M^{-1}$, so $M^\dagger$ is a "generalized inverse" of M. Even though the definition of $M^\dagger$ seems to depend on U and V, actually, $M^\dagger$ is uniquely defined in terms of M (the same $M^\dagger$ is obtained for all possible SVD decompositions of M). It is easy to check that

$$\begin{aligned} MM^\dagger M &= M \\ M^\dagger MM^\dagger &= M^\dagger \end{aligned}$$

and both $MM^\dagger$ and $M^\dagger M$ are symmetric matrices. In fact,

$$MM^\dagger = U\Sigma V^\top V\Sigma^\dagger U^\top = U\Sigma\Sigma^\dagger U^\top = U\begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} U^\top$$

and

$$M^\dagger M = V\Sigma^\dagger U^\top U\Sigma V^\top = V\Sigma^\dagger\Sigma V^\top = V\begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} V^\top.$$

We immediately get

$$\begin{aligned} (MM^\dagger)^2 &= MM^\dagger \\ (M^\dagger M)^2 &= M^\dagger M, \end{aligned}$$

so both $MM^\dagger$ and $M^\dagger M$ are orthogonal projections (since they are both symmetric). We claim that $MM^\dagger$ is the orthogonal projection onto the range of M and $M^\dagger M$ is the orthogonal projection onto $\mathrm{Ker}(M)^\perp$, the orthogonal complement of $\mathrm{Ker}(M)$.

Obviously, $\mathrm{range}(MM^\dagger) \subseteq \mathrm{range}(M)$ and for any $y = Mx \in \mathrm{range}(M)$, as $MM^\dagger M = M$, we have

$$MM^\dagger y = MM^\dagger Mx = Mx = y,$$

so the image of $MM^\dagger$ is indeed the range of M. It is also clear that $\mathrm{Ker}(M) \subseteq \mathrm{Ker}(M^\dagger M)$ and since $MM^\dagger M = M$, we also have $\mathrm{Ker}(M^\dagger M) \subseteq \mathrm{Ker}(M)$ and so,

$$\mathrm{Ker}(M^\dagger M) = \mathrm{Ker}(M).$$

Since $M^\dagger M$ is Hermitian, $\mathrm{range}(M^\dagger M) = \mathrm{Ker}(M^\dagger M)^\perp = \mathrm{Ker}(M)^\perp$, as claimed.

It will also be useful to see that $\mathrm{range}(M) = \mathrm{range}(MM^\dagger)$ consists of all vector $y \in \mathbb{R}^n$ such that

$$U^\top y = \begin{pmatrix} z \\ 0 \end{pmatrix},$$

with $z \in \mathbb{R}^r$.

Indeed, if $y = Mx$, then

$$U^\top y = U^\top Mx = U^\top U \Sigma V^\top x = \Sigma V^\top x = \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} V^\top x = \begin{pmatrix} z \\ 0 \end{pmatrix},$$

where Σ_r is the $r \times r$ diagonal matrix $\mathrm{diag}(\sigma_1, \ldots, \sigma_r)$. Conversely, if $U^\top y = \binom{z}{0}$, then $y = U\binom{z}{0}$ and

$$\begin{aligned} MM^\dagger y &= U \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} U^\top y \\ &= U \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} U^\top U \begin{pmatrix} z \\ 0 \end{pmatrix} \\ &= U \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} \begin{pmatrix} z \\ 0 \end{pmatrix} \\ &= U \begin{pmatrix} z \\ 0 \end{pmatrix} = y, \end{aligned}$$

which shows that y belongs to the range of M.

Similarly, we claim that $\mathrm{range}(M^\dagger M) = \mathrm{Ker}(M)^\perp$ consists of all vector $y \in \mathbb{R}^n$ such that

$$V^\top y = \begin{pmatrix} z \\ 0 \end{pmatrix},$$

with $z \in \mathbb{R}^r$.

If $y = M^\dagger M u$, then

$$y = M^\dagger M u = V \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} V^\top u = V \begin{pmatrix} z \\ 0 \end{pmatrix},$$

for some $z \in \mathbb{R}^r$. Conversely, if $V^\top y = \binom{z}{0}$, then $y = V\binom{z}{0}$ and so,

$$\begin{aligned} M^\dagger M V \begin{pmatrix} z \\ 0 \end{pmatrix} &= V \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} V^\top V \begin{pmatrix} z \\ 0 \end{pmatrix} \\ &= V \begin{pmatrix} I_r & 0 \\ 0 & 0_{n-r} \end{pmatrix} \begin{pmatrix} z \\ 0 \end{pmatrix} \\ &= V \begin{pmatrix} z \\ 0 \end{pmatrix} = y, \end{aligned}$$

which shows that $y \in \mathrm{range}(M^\dagger M)$.

If M is a symmetric matrix, then in general, there is no SVD, $U\Sigma V^\top$, of M with $U = V$. However, if $M \succeq 0$, then the eigenvalues of M are nonnegative and so the nonzero eigenvalues of M are equal to the singular values of M and SVD's of M are of the form

$$M = U\Sigma U^\top.$$

Analogous results hold for complex matrices but in this case, U and V are unitary matrices and $MM^\dagger$ and $M^\dagger M$ are Hermitian orthogonal projections.

If M is a normal matrix which, means that $MM^\top = M^\top M$, then there is an intimate relationship between SVD's of M and block diagonalizations of M. As a consequence, the pseudo-inverse of a normal matrix, M, can be obtained directly from a block diagonalization of M.

If M is a (real) normal matrix, then it can be block diagonalized with respect to an orthogonal matrix, U, as

$$M = U\Lambda U^\top,$$

where Λ is the (real) block diagonal matrix,

$$\Lambda = \mathrm{diag}(B_1, \ldots, B_n),$$

consisting either of 2×2 blocks of the form

$$B_j = \begin{pmatrix} \lambda_j & -\mu_j \\ \mu_j & \lambda_j \end{pmatrix}$$

with $\mu_j \neq 0$, or of one-dimensional blocks, $B_k = (\lambda_k)$. Assume that $B_1, \ldots, B_p$ are 2×2 blocks and that $\lambda_{2p+1}, \ldots, \lambda_n$ are the scalar entries. We know that the numbers $\lambda_j \pm i\mu_j$, and the λ_{2p+k} are the eigenvalues of A. Let $\rho_{2j-1} = \rho_{2j} = \sqrt{\lambda_j^2 + \mu_j^2}$ for $j = 1, \ldots, p$, $\rho_{2p+j} = \lambda_j$ for $j = 1, \ldots, n - 2p$, and assume that the blocks are ordered so that $\rho_1 \geq \rho_2 \geq \cdots \geq \rho_n$. Then, it is easy to see that

$$UU^\top = U^\top U = U\Lambda U^\top U\Lambda^\top U^\top = U\Lambda\Lambda^\top U^\top,$$

with

$$\Lambda\Lambda^\top = \mathrm{diag}(\rho_1^2, \ldots, \rho_n^2)$$

so, the singular values, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, of A, which are the nonnegative square roots of the eigenvalues of $AA^\top$, are such that

$$\sigma_j = \rho_j, \quad 1 \leq j \leq n.$$

We can define the diagonal matrices

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_r, 0, \ldots, 0)$$

where $r = \mathrm{rank}(A)$, $\sigma_1 \geq \cdots \geq \sigma_r > 0$, and

$$\Theta = \mathrm{diag}(\sigma_1^{-1} B_1, \ldots, \sigma_{2p}^{-1} B_p, 1, \ldots, 1),$$

so that Θ is an orthogonal matrix and

$$\Lambda = \Theta\Sigma = (B_1, \ldots, B_p, \lambda_{2p+1}, \ldots, \lambda_r, 0, \ldots, 0).$$

But then, we can write

$$A = U\Lambda U^\top = U\Theta\Sigma U^\top$$

and we if let $V = U\Theta$, as U is orthogonal and Θ is also orthogonal, V is also orthogonal and $A = V\Sigma U^\top$ *is an SVD for* A. Now, we get

$$A^+ = U\Sigma^+ V^\top = U\Sigma^+\Theta^\top U^\top.$$

However, since Θ is an orthogonal matrix, $\Theta^\top = \Theta^{-1}$ and a simple calculation shows that

$$\Sigma^+\Theta^\top = \Sigma^+\Theta^{-1} = \Lambda^+,$$

which yields the formula

$$A^+ = U\Lambda^+ U^\top.$$

Also observe that if we write

$$\Lambda_r = (B_1, \ldots, B_p, \lambda_{2p+1}, \ldots, \lambda_r),$$

then Λ_r is invertible and

$$\Lambda^+ = \begin{pmatrix} \Lambda_r^{-1} & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore, the pseudo-inverse of a normal matrix can be computed directly from any block diagonalization of A, as claimed.

Next, we will use pseudo-inverses to generalize the result of Section 2 to symmetric matrices $M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$ where C (or A) is singular.

4 A Characterization of Symmetric Positive Semidefinite Matrices Using Schur Complements

We begin with the following simple fact:

Proposition 4.1 *If P is an invertible symmetric matrix, then the function*

$$f(x) = \frac{1}{2}x^\top Px + x^\top b$$

has a minimum value iff $P \succeq 0$, in which case this optimal value is obtained for a unique value of x, namely $x^* = -P^{-1}b$, and with

$$f(P^{-1}b) = -\frac{1}{2}b^\top P^{-1}b.$$

Proof. Observe that

$$\frac{1}{2}(x + P^{-1}b)^\top P(x + P^{-1}b) = \frac{1}{2}x^\top Px + x^\top b + \frac{1}{2}b^\top P^{-1}b.$$

Thus,

$$f(x) = \frac{1}{2}x^\top Px + x^\top b = \frac{1}{2}(x + P^{-1}b)^\top P(x + P^{-1}b) - \frac{1}{2}b^\top P^{-1}b.$$

If P has some negative eigenvalue, say $-\lambda$ (with $\lambda > 0$), if we pick any eigenvector, u, of P associated with λ, then for any $\alpha \in \mathbb{R}$ with $\alpha \neq 0$, if we let $x = \alpha u - P^{-1}b$, then as $Pu = -\lambda u$ we get

$$\begin{aligned} f(x) &= \frac{1}{2}(x + P^{-1}b)^\top P(x + P^{-1}b) - \frac{1}{2}b^\top P^{-1}b \\ &= \frac{1}{2}\alpha u^\top P\alpha u - \frac{1}{2}b^\top P^{-1}b \\ &= -\frac{1}{2}\alpha^2\lambda \left\|u\right\|_2^2 - \frac{1}{2}b^\top P^{-1}b, \end{aligned}$$

and as α can be made as large as we want and $\lambda > 0$, we see that f has no minimum. Consequently, in order for f to have a minimum, we must have $P \succeq 0$. In this case, as $(x + P^{-1}b)^\top P(x + P^{-1}b) \geq 0$, it is clear that the minimum value of f is achieved when $x + P^{-1}b = 0$, that is, $x = -P^{-1}b$. $\square$

Let us now consider the case of an arbitrary symmetric matrix, P.

Proposition 4.2 *If P is a symmetric matrix, then the function*

$$f(x) = \frac{1}{2}x^\top Px + x^\top b$$

has a minimum value iff $P \succeq 0$ and $(I - PP^\dagger)b = 0$, in which case this minimum value is

$$p^* = -\frac{1}{2} b^\top P^\dagger b.$$

Furthermore, if $P = U^\top \Sigma U$ is an SVD of P, then the optimal value is achieved by all $x \in \mathbb{R}^n$ of the form

$$x = -P^\dagger b + U^\top \begin{pmatrix} 0 \\ z \end{pmatrix},$$

for any $z \in \mathbb{R}^{n-r}$, where r is the rank of P.

Proof. The case where P is invertible is taken care of by Proposition 4.1 so, we may assume that P is singular. If P has rank $r < n$, then we can diagonalize P as

$$P = U^\top \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} U,$$

where U is an orthogonal matrix and where Σ_r is an $r \times r$ diagonal invertible matrix. Then, we have

$$\begin{aligned} f(x) &= \frac{1}{2} x^\top U^\top \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} U x + x^\top U^\top U b \\ &= \frac{1}{2} (Ux)^\top \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} U x + (Ux)^\top U b. \end{aligned}$$

If we write $Ux = \binom{y}{z}$ and $Ub = \binom{c}{d}$, with $y, c \in \mathbb{R}^r$ and $z, d \in \mathbb{R}^{n-r}$, we get

$$\begin{aligned} f(x) &= \frac{1}{2} (Ux)^\top \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} U x + (Ux)^\top U b \\ &= \frac{1}{2} (y^\top, z^\top) \begin{pmatrix} \Sigma_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} y \\ z \end{pmatrix} + (y^\top, z^\top) \begin{pmatrix} c \\ d \end{pmatrix} \\ &= \frac{1}{2} y^\top \Sigma_r y + y^\top c + z^\top d. \end{aligned}$$

For $y = 0$, we get

$$f(x) = z^\top d,$$

so if $d \neq 0$, the function f has no minimum. Therefore, if f has a minimum, then $d = 0$. However, $d = 0$ means that $Ub = \binom{c}{0}$ and we know from Section 3 that b is in the range of P (here, U is $U^\top$) which is equivalent to $(I - PP^\dagger)b = 0$. If $d = 0$, then

$$f(x) = \frac{1}{2} y^\top \Sigma_r y + y^\top c$$

and as Σ_r is invertible, by Proposition 4.1, the function f has a minimum iff $\Sigma_r \succeq 0$, which is equivalent to $P \succeq 0$.

Therefore, we proved that if f has a minimum, then $(I - PP^\dagger)b = 0$ and $P \succeq 0$. Conversely, if $(I - PP^\dagger)b = 0$ and $P \succeq 0$, what we just did proves that f does have a minimum.

When the above conditions hold, the minimum is achieved if $y = -\Sigma_r^{-1}c$, $z = 0$ and $d = 0$, that is for x^* given by $Ux^* = \binom{-\Sigma_r^{-1}c}{0}$ and $Ub = \binom{c}{0}$, from which we deduce that

$$x^* = -U^\top \begin{pmatrix} \Sigma_r^{-1}c \\ 0 \end{pmatrix} = -U^\top \begin{pmatrix} \Sigma_r^{-1}c & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} c \\ 0 \end{pmatrix} = -U^\top \begin{pmatrix} \Sigma_r^{-1}c & 0 \\ 0 & 0 \end{pmatrix} Ub = -P^\dagger b$$

and the minimum value of f is

$$f(x^*) = -\frac{1}{2} b^\top P^\dagger b.$$

For any $x \in \mathbb{R}^n$ of the form

$$x = -P^\dagger b + U^\top \begin{pmatrix} 0 \\ z \end{pmatrix}$$

for any $z \in \mathbb{R}^{n-r}$, our previous calculations show that $f(x) = -\frac{1}{2} b^\top P^\dagger b$. $\square$

We now return to our original problem, characterizing when a symmetric matrix, $M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$, is positive semidefinite. Thus, we want to know when the function

$$f(x, y) = (x^\top, y^\top) \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = x^\top A x + 2x^\top B y + y^\top C y$$

has a minimum with respect to both x and y. Holding y constant, Proposition 4.2 implies that $f(x, y)$ has a minimum iff $A \succeq 0$ and $(I - AA^\dagger)By = 0$ and then, the minimum value is

$$f(x^*, y) = -y^\top B^\top A^\dagger B y + y^\top C y = y^\top (C - B^\top A^\dagger B) y.$$

Since we want $f(x, y)$ to be uniformly bounded from below for all x, y, we must have $(I - AA^\dagger)B = 0$. Now, $f(x^*, y)$ has a minimum iff $C - B^\top A^\dagger B \succeq 0$. Therefore, we established that $f(x, y)$ has a minimum over all x, y iff

$$A \succeq 0, \quad (I - AA^\dagger)B = 0, \quad C - B^\top A^\dagger B \succeq 0.$$

A similar reasoning applies if we first minimize with respect to y and then with respect to x but this time, the Schur complement, $A - BC^\dagger B^\top$, of C is involved. Putting all these facts together we get our main result:

Theorem 4.3 *Given any symmetric matrix, $M = \begin{pmatrix} A & B \\ B^\top & C \end{pmatrix}$, the following conditions are equivalent:*

(1) $M \succeq 0$ (M is positive semidefinite).

(2) $A \succeq 0, \quad (I - AA^{\dagger})B = 0, \quad C - B^{\top}A^{\dagger}B \succeq 0.$

(2) $C \succeq 0, \quad (I - CC^{\dagger})B^{\top} = 0, \quad A - BC^{\dagger}B^{\top} \succeq 0.$

If $M \succeq 0$ as in Theorem 4.3, then it is easy to check that we have the following factorizations (using the fact that $A^{\dagger}AA^{\dagger} = A^{\dagger}$ and $C^{\dagger}CC^{\dagger} = C^{\dagger}$):

$$\begin{pmatrix} A & B \\ B^{\top} & C \end{pmatrix} = \begin{pmatrix} I & BC^{\dagger} \\ 0 & I \end{pmatrix} \begin{pmatrix} A - BC^{\dagger}B^{\top} & 0 \\ 0 & C \end{pmatrix} \begin{pmatrix} I & 0 \\ C^{\dagger}B^{\top} & I \end{pmatrix}$$

and

$$\begin{pmatrix} A & B \\ B^{\top} & C \end{pmatrix} = \begin{pmatrix} I & 0 \\ B^{\top}A^{\dagger} & I \end{pmatrix} \begin{pmatrix} A & 0 \\ 0 & C - B^{\top}A^{\dagger}B \end{pmatrix} \begin{pmatrix} I & A^{\dagger}B \\ 0 & I \end{pmatrix}.$$

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, first edition, 2004.

[2] James W. Demmel. *Applied Numerical Linear Algebra*. SIAM Publications, first edition, 1997.

[3] Jean H. Gallier. *Geometric Methods and Applications, For Computer Science and Engineering*. TAM, Vol. 38. Springer, first edition, 2000.

[4] H. Golub, Gene and F. Van Loan, Charles. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.

[5] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, first edition, 1990.

[6] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, first edition, 1994.

[7] Gilbert Strang. *Linear Algebra and its Applications*. Saunders HBJ, third edition, 1988.

[8] L.N. Trefethen and D. Bau III. *Numerical Linear Algebra*. SIAM Publications, first edition, 1997.