Affinely Constrained Rayleigh Quotients

Timothee Cour

We present an efficient method to maximize Generalized Rayleigh Quotients under affine constraints (ACRQ). We show it can be solved by an eigendecomposition, which leads to more efficient solvers than an equivalent QCQP formulation.

1 Affinely Constrained Generalized Rayleigh Quotients

The general formulation of the Affinely Constrained Generalized Rayleigh Quotient (ACRQ) is as follows:

$$\text{Maximize } \epsilon(x) = \frac{x^T A x}{x^T B x} \tag{1}$$

$$\text{Subject to } Cx = d \tag{2}$$

where $x \in \mathbb{R}^n$, $b \in \mathbb{R}^k$, $A, B \in S^n$, $B > 0$, and $C \in R^{k \times n}$ is full-rank, k ($k < n$).

1.1 Special Case: $B = I_n$, $d = 0$ (*Linear* Constraint)

The solution to this problem has been proposed in [1, 2]. We give here a summary. After computing the Lagrangian, we obtain the following unconstrained Rayleigh Quotient Maximization:

$$\text{Maximize } \epsilon(x) = \frac{x^T PAPx}{x^T x}, \tag{3}$$

where we have introduced the feasible subspace projection matrix

$$P = I_n - C^T(CC^T)^{-1}C \tag{4}$$

The solution of the program can be computed using the leading eigenvector of PAP.

1.2 General Case: how to reformulate the problem

We assume here that $b \neq 0$, otherwise we use the previous section. First we get rid of the matrix B at the denominator. With the change of variable $x = B^{-1/2}y$ (which exists [1] since $B > 0$) we obtain:

$$\text{Maximize } \epsilon_1(y) = \frac{y^T A' y}{y^T y} \tag{5}$$

$$\text{Subject to } C'y = d, \tag{6}$$

with $A' = B^{-1/2}AB^{-1/2}$ and $C' = CB^{-1/2}$.

Second, we notice that (5) is equivalent to:

$$\text{Maximize } \epsilon_2(z, t) = \frac{z^T A' z}{z^T z} \tag{7}$$

$$\text{Subject to } C'z = td, \tag{8}$$

where we introduced the new variable $t \in \mathbb{R}$. Indeed, (z^*, t^*) is an optimum of (7) *iff* $y^* = (1/t^*)z^*$ is an optimum of (5). The special case $t^* = 0$ will be taken care of in section 1.3. We now explore three different approaches to solve (7), and explain their pros and cons. The first one increases the number of variables by 1, while the other two decrease the number of constraints by 1.

Solution 1. The most natural and common solution is to increase the dimension by 1, transforming the affine constraint into a linear one:

$$\text{Maximize } \epsilon_3(\bar{z}) = \frac{\bar{z}^T \bar{A} \bar{z}}{\bar{z}^T \bar{I_n} \bar{z}} \tag{9}$$

$$\text{Subject to } \bar{C}\bar{z} = 0, \tag{10}$$

where $\bar{z} = \begin{bmatrix} z \\ t \end{bmatrix}$, $\bar{A} = \begin{bmatrix} A' & 0 \\ 0 & 0 \end{bmatrix}$, $\bar{C} = \begin{bmatrix} C' & -b \end{bmatrix}$, and $\bar{I_n} = \begin{bmatrix} I_n & 0 \\ 0 & 0 \end{bmatrix}$. The problem is that the denominator matrix does not satisfy $\bar{I_n} > 0$ anymore, so we

[1] In many practical problems, the computation of $B^{-1/2}$ is not a bottleneck, either because B is a diagonal matrix or because its Cholesky factorization is simple. In the latter case we never compute the matrix $B^{-1/2}$ but instead we solve triangular systems on the fly

cannot fall back into the linear case treated in section 1.1. Note, the problem can be cast as a Quadratically Constrained Quadratic Program (QCQP):

$$\text{Maximize } \epsilon_4(\bar{z}') = \bar{z}'^T \bar{A} \bar{z}' \quad (11)$$

$$\text{Subject to } \bar{C}\bar{z}' = 0,\ \bar{z}'^T \bar{I}_n \bar{z}' \leq 1, \quad (12)$$

but this is not very useful and can be dealt with more efficiently as we shall see.

Solution 2. Notice that $[\exists t : C'z = tb] \Leftrightarrow C'z \in Span\{b\} \Leftrightarrow C'z \in \ker K_b^T$, where K_b is a matrix whose range is $b^\perp$. The matrix $K_b = I_k - bb^T/||b||^2 \in \mathbb{R}^{k\times k}$ does the job, since we have: $K_b^T u = 0 \Rightarrow u \in Span\{b\}$ and $K_b^T b = 0$, so we have the equivalent problem:

$$\text{Maximize } \epsilon_5(z) = \frac{z^T A' z}{z^T z} \quad (13)$$

$$\text{Subject to } K_b C' z = 0, \quad (14)$$

since K_b is symmetric. There is one last detail: $K_b C'$ is not full rank because $rank(K_b) = k - 1$. Assume WLOG that $b_k \neq 0$ (otherwise reorder the rows of C' and b). Letting $J = \begin{bmatrix} I_{k-1} & 0 \end{bmatrix}$ be the canonical projector $\mathbb{R}^k \to \mathbb{R}^{k-1}$, simple calculus shows that $C_{eq} = JK_bC'$ is a full rank matrix with same kernel as K_bC'. We now obtain a valid linearly constrained Rayleigh Quotient in the form of:

$$\text{Maximize } \epsilon_6(z) = \frac{z^T A' z}{z^T z} \quad (15)$$

$$\text{Subject to } JK_bC'z = 0 \quad (16)$$

In that form , we can solve this problem with the results of section 1.1.

Extension. The solution outlined here can be readily generalized to handle the following constraint:

$$C'z \in Span\{b_1, ..., b_s\} \quad (17)$$

(before we had $s = 1$). In this case, letting $B = [b_1...b_s]$, we take $K_B = I_k - B(B^TB)^{-1}B^T$ and $J \in \mathbb{R}^{k-s\times k}$, that selects $k - s$ independant rows of K_B.

Solution 3. The solution 2 above is workable, and may be the only way to deal with the aforementioned extension, but it may not be the most efficient method if the matrices are structured. The third solution takes advantage of any structure in the constraint matrix.

Assuming $b_k \neq 0$ as before, $[\exists t : C'z = tb] \Leftrightarrow \forall i \in [1, k-1],\ (C'z)_i = (1/b_k)(C'z)_k b_i \Leftrightarrow J(C' - (1/b_k)bC'_k)z = 0$, with C'_k the last row of C. We can therefore take $C_{eq} = J(C' - (1/b_k)bC'_k)$ as our new equivalent linear constraint matrix. It has the same kernel and therefore is also full rank.

Comparison. Recall that the solution to the linear case 1.1 requires computing the inverse $(C_{eq}C_{eq}^T)^{-1}$ where C_{eq} is the matrix of the equivalent linear constraint formulation. Assume C is sparse and b is a full $k \times 1$ vector. In solution 2, $C_{eq} = J(I_k - bb^T/||b||^2)C'$ is full because of bb^T, which could be a problem for large k. However, it is easy to show that in solution 3, $C_{eq} = J(C' - (1/b_k)bC'_k)$ is as sparse as C'. We can therefore compute efficiently a sparse Cholesky factorization of $C_{eq}C_{eq}^T$ and compute $y = (C_{eq}C_{eq}^T)^{-1}x$ via two triangular solves.

1.3 Special case $t^* = 0$ in (7)

In that case, the original problem (1) has no solution, even though (7) has one. In this case, the leading eigenvector of A' and the kernel o C' are parallel, and only a diverging sequence of points approximates the supremum of (1).

2 Appendix

See the main paper for notations.

Proposition 2.1 **($X_{orth} \in \Omega$)** *When $n = n'$, we have $X \in \Omega \Rightarrow X_{orth} \in \Omega$. More generally, $\forall A \in \mathbb{R}^{n\times n}$, whenever $u \in \mathbb{R}^n$ is a left* **and** *right eigenvector of A, then u is also a left and right eigenvector of A_{orth}.*

Note that in general, X and X_{orth} do *not* have the same eigenvectors, here we are lucky because of the particular constraint induced by Ω.

Proof Since u is eigenvector of A *and* A^T, we can show that u is a left and right singular vector of A (i.e. it is a column in U and in V with $A = U\Sigma V^T$) and therefore, all other singular vectors in U and V are orthogonal to u (although X is not necessarily symmetric). Therefore, $UV^Tu = cst \cdot U[1, 0...0]^T = cst' \cdot u \Rightarrow X_{orth}u = cst' \cdot u$. Same with X_{orth}^T. Now we can show that $X_{orth} \in \Omega$. Since $X \in \Omega$, taking $u = \mathbf{1}$, we have $X_{orth}\mathbf{1} = cst'\mathbf{1}$, but because $X_{orth} \in O(n)$, we have $|cst'| = 1$, and it is easy to show that the eigenvalue associated with u in A and A_{orth} have same sign, so we conclude here that $X_{orth}\mathbf{1} = \mathbf{1}$, and likewise $X_{orth}^T\mathbf{1} = \mathbf{1}\square$

References

[1] Golub G.H. von Matt U. Gander, W. A constrained eigenvalue problem. In *Linear Algebra Appl. 114-115, pp. 815-839*, 1989.

[2] Stella X. Yu and Jianbo Shi. Grouping with bias. In *Advances in Neural Information Processing Systems*, 2001.