Supplemental materials for “Penalized versus constrained generalized eigenvalue problems”

Irina Gaynanova, James G. Booth and Martin T. Wells

March 17, 2016

S1 Derivation of Algorithm 1.

Following Witten & Tibshirani (2011), (5) can be recast as a biconvex optimization problem

$$\text{maximize}_{u,v}\left\{2u^\top Q^{1/2}v-\lambda\sum_{j=1}^p|v_j|-u^\top u\right\}\text{subject to } v^\top v\leq 1, \tag{S1}$$

since maximizing with respect to u gives $u=Q^{1/2}v$. The problem (S1) is convex with respect to u when v is fixed and is convex with respect to v when u is fixed. This property allows the use of Alternate Convex Search (ACS) to find the solution (Gorski et al. 2007, Section 4.2.1). ACS ensures that all accumulation points are partial optima and have the same function value (Gorski et al. 2007, Theorem 4.9).

Starting with an initial value $v^{(0)}$ the algorithm proceeds by iterating the following two steps:

Step 1 $u^{(k)}=\arg\max_u\left\{2u^\top Q^{1/2}v^{(k)}-u^\top u\right\}=Q^{1/2}v^{(k)}$

Step 2 $v^{(k+1)}=\arg\max_v\left\{2(u^{(k)})^\top Q^{1/2}v-\lambda\sum_{j=1}^p|v_j|\right\}$ subject to $v^\top v\leq 1$.

Following Witten & Tibshirani (2011, Proposition 2), it is useful to reformulate Step 2 as

$$q^{(k+1)}=\arg\max_q\left\{2(u^{(k)})^\top Q^{1/2}q-\lambda\sum_{j=1}^p|q_j|-q^\top q\right\} \tag{S2}$$

where, if $q^{(k+1)}=0$, then $v^{(k+1)}=0$, else $v^{(k+1)}=q^{(k+1)}/\sqrt{(q^{(k+1)})^\top q^{(k+1)}}$. Since problem (S2) is convex with respect to q, the solution $q^{(k+1)}$ satisfies KKT conditions (Boyd & Vandenberghe 2004)

$$2Q^{1/2}u^{(k)}-2q^{(k+1)}-\lambda\Gamma=0, \tag{S3}$$

where Γ is a p-vector and each Γ_j is a subgradient of $|q_j^{(k+1)}|$, i.e. $\Gamma_j=1$ if $q_j^{(k+1)}>0$, $\Gamma_j=-1$ if $q_j^{(k+1)}<0$ and Γ_j is between -1 and 1 if $q_j^{(k+1)}=0$. From (S3)

$$q_j^{(k+1)}=\operatorname{sign}((Q^{1/2}u^{(k)})_j)\left(|(Q^{1/2}u^{(k)})_j|-\frac{\lambda}{2}\right)_+. \tag{S4}$$

Algorithm 1 results from combining Steps 1 and 2 with the update (S4).

S2 Proofs.

Proof of Proposition 1. Let q_i be the ith row of Q and $q_i(k)$ be the subvector of q_i of length k with the maximal ℓ_2 norm. If v satisfies $\|v\|_0 \le k$, then $|q_i^\top v| \le \|q_i(k)\|_2 \|v\|_2$. Furthermore, if $v^\top C v \le 1$, then $\|v\|_2 = \|C^{-1/2} C^{1/2} v\|_2 \le \|C^{-1/2}\| \|C^{1/2} v\|_2 \le \|C^{-1/2}\|$, where $\|C^{-1/2}\|$ is the spectral norm of matrix $C^{-1/2}$. Let $\sigma_{\min}(C)$ be the smallest eigenvalue of C. It follows that for all v such that $\|v\|_0 \le k$ and $v^\top C v \le 1$,

$$
\begin{aligned}
v^\top Q v - \lambda \|v\|_1 &= \sum_{i=1}^p |v_i| \left(\operatorname{sign}(v_i) q_i^\top v - \lambda\right) \\
&\le \sum_{i=1}^p |v_i| (\|q_i(k)\|_2 \|v\|_2 - \lambda) \\
&\le \sum_{i=1}^p |v_i| \left(\frac{\|q_i(k)\|_2}{\sqrt{\sigma_{\min}(C)}} - \lambda \right).
\end{aligned}
$$

Let $\tilde{q} \in \mathbb{R}^p$ be a vector with elements $\tilde{q}_i = \left(\|q_i(k)\|_2 - \lambda \sqrt{\sigma_{\min}(C)}\right)_+$, where $x_+ = \max(x, 0)$, and let $\tilde{q}(k)$ be the subvector of $\tilde{q}$ of length k with the maximal ℓ_2 norm. From above, for all v such that $\|v\|_0 \le k$ and $v^\top C v \le 1$

$$
\begin{aligned}
v^\top Q v - \lambda \|v\|_1 &\le \frac{1}{\sqrt{\sigma_{\min}(C)}} \sum_{i=1}^p |v_i| \tilde{q}_i \\
&\le \frac{1}{\sqrt{\sigma_{\min}(C)}} \|\tilde{q}(k)\|_2 \|v\|_2 \\
&\le \frac{1}{\sigma_{\min}(C)} \|\tilde{q}(k)\|_2.
\end{aligned}
$$

□

Proof of Proposition 2. Fix any $\lambda \ge 0$ and let v_λ be the solution to (4). It follows that for any v such that $v^\top v \le 1$,

$$
v_\lambda^\top Q v_\lambda - \lambda \|v_\lambda\|_1 \ge v^\top Q v - \lambda \|v\|_1. \tag{S5}
$$

Consider (6) with $\tau = \|v_\lambda\|_1$. From (S5), for each v such that $v^\top v \le 1$ and $\|v\|_1 \le \tau$,

$$
v_\lambda^\top Q v_\lambda \ge v^\top Q v + \lambda(\|v_\lambda\|_1 - \|v\|_1) = v^\top Q v + \lambda(\tau - \|v\|_1) \ge v^\top Q v.
$$

This means v_λ is the solution to (6), hence $v_\tau = v_\lambda$. □

Proof of Proposition 3. The result for $\tau_{\max}$ follows from the fact that $v^{(0)}$ is the solution to the unconstrained GEP:

$$
\begin{aligned}
v^{(0)} = \arg\max_v \quad & v^\top Q v \\
\text{s.t.} \quad & v^\top C v \le 1.
\end{aligned}
$$

The result for $\tau = 1$ follows from $v^\top Qv \leq \|v\|_1^2 \max_{ij}|q_{ij}| \leq \max_i q_{ii}$. If $q_{kk} = \max_i q_{ii}$, then the upper bound is obtained when $v_k = 1$ and $v_i = 0$ for $i \neq k$. Since this v satisfies the constraints $v^\top Cv \leq 1$ and $\|v\|_1 \leq 1$, it follows that $\|v_\tau\|_0 = 1$.

The result for $\tau > 1$ follows from $\|v\|_1 > 1$ and $v^\top v \leq 1$ implying $\|v\|_0 \geq 2$.

□

Lemma 1. *Define $T(\lambda)$ as*

$$T(\lambda) = \frac{\sum_{i=1}^p (|a_i| - \lambda)_+}{\sqrt{\sum_{i=1}^p (|a_i| - \lambda)_+^2}},$$

where $a \in \mathbb{R}^p$ has elements $|a_1| \geq |a_2| \geq ... \geq |a_p| > 0$ and $\lambda \in [0, |a_1|)$. Then $T(\lambda)$ is a decreasing function of λ.

Proof. Consider the strict inequality $|a_s| > |a_{s+1}|$ for some $s = 1, ..., p$ (with the convention $a_{p+1} = 0$) and let $\lambda \in [|a_{s+1}|, |a_s|]$. For such λ, by definition

$$T(\lambda) = \frac{\sum_{i=1}^s (|a_i| - \lambda)}{\sqrt{\sum_{i=1}^s (|a_i| - \lambda)^2}}.$$

Since $T'(\lambda) < 0$, it follows that $T(\lambda)$ is decreasing on $\lambda \in [|a_s|, |a_{s+1}|]$. By continuity, it follows that $T(\lambda)$ is decreasing for all λ satisfying $\max_i |a_i| > \lambda \geq 0$. □

Proof of Proposition 4. Since Q is rank one, it follows that maximization of $v^\top Qv$ is equivalent to the maximization of $l^\top v$, where l is the eigenvector of Q. Hence,

$$\begin{aligned} v_\tau = \arg\max_v \quad & \{l^\top v\} \\ \text{s.t.} \quad & v^\top v \leq 1 \\ & \|v\|_1 \leq \tau. \end{aligned}$$

From KKT conditions, the solution v_τ has components $v_{\tau,j}$, $i = 1, ..., p$, where

$$v_{\tau,j} = \frac{(|l_j| - \lambda)_+}{\sqrt{\sum_{i=1}^p (|l_i| - \lambda)_+^2}},$$

and $\lambda \geq 0$ is such that $\|v_\tau\|_1 = \tau$. Moreover, if $\lambda \in (|l_s|, |l_s + 1|]$, then v_τ has exactly s nonzero components. Let

$$T(\lambda) = \frac{\sum_{i=1}^p (|l_i| - \lambda)_+}{\sqrt{\sum_{i=1}^p (|l_i| - \lambda)_+^2}}.$$

Then $T(\lambda)$ is a continuous and decreasing function of λ for all $0 \leq \lambda < |l_1|$ (Lemma 1). Therefore, for every $\tilde{\lambda} \in (|l_s|, |l_s + 1|]$, there exists $\tau \geq 1$ such that $T(\tilde{\lambda}) = \tau$. Since $T(\lambda)$ is strictly decreasing, it means that there exists τ such that $T^{-1}(\tau) \in (|l_s|, |l_s + 1|]$, and as a result $\|v_\tau\|_0 = s$. □

Assumption 1. *Matrices Q and C are such that for any $s \in \{2, ..., p-1\}$*

$$\begin{array}{lll} \max\limits_{v} & v^\top Qv \\ s.t. & v^\top Cv \leq 1 \\ & \|v\|_0 = s-1 \end{array} < \begin{array}{lll} \max\limits_{v} & v^\top Qv \\ s.t. & v^\top Cv \leq 1 \\ & \|v\|_0 = s \end{array} < \begin{array}{lll} \max & v^\top Qv \\ s.t. & v^\top Cv \leq 1 \\ & \|v\|_0 = s+1 \end{array}$$

This assumption states that different sparsity levels lead to different values of $v^\top Qv$. When Q is rank one, and $C = I$, it reduces to the conditions of Proposition 4.

Assumption 2. *If τ is such that*

$$\begin{array}{lll} \max & v^\top Qv \\ s.t. & v^\top Cv \leq 1 \\ & \|v\|_1 \leq \tau \end{array} < \begin{array}{lll} \max\limits_{v} & v^\top Qv \\ s.t. & v^\top Cv \leq 1 \\ & \|v\|_0 = k, \end{array}$$

then $\|v_\tau\|_0 \leq k + t_k$, where $0 \leq t_k \leq p - k$.

This assumption states that in case the restriction on ℓ_1 norm eliminates the best k-sparse solution, the resulting v_τ can have at most $k + t_k$ non-zero elements. It is always satisfied for $t_k = p - k$.

Proof of Proposition 5. For every $s \in \{1, ..., p\}$, define

$$\begin{array}{rll} v_s = \arg\max\limits_{v} & v^\top Qv \\ \text{s.t.} & v^\top Cv \leq 1 \\ & \|v\|_0 = s \end{array}$$

Let $q_s = v_s^\top Qv_s$ and $\tau_s = \|v_s\|_1$. From Proposition 3, $\tau_1 = 1$ and $\tau_p = \tau_{\max}$. Since $v_\tau^\top Qv_\tau$ is nondecreasing and continuous in τ, and $0 < q_1 < ... < q_p$ by Assumption 1, for every $s \in \{2, ..., p\}$, there exists $\tau \in (1, \tau_{\max})$ such that $q_{s-1} < v_\tau^\top Qv_\tau < q_s$. The first inequality, $q_{s-1} < v_\tau^\top Qv_\tau$, implies that $\|v_\tau\|_0 \geq s$ (as q_{s-1} is the maximal value obtained when $\|v\|_0 \leq s-1$). The second inequality implies $\|v_\tau\|_0 \leq s + t_s$ by Assumption 2. $\square$

References

Boyd, S. P. & Vandenberghe, L. (2004), *Convex Optimization*, Cambridge Univ Press, Cambridge.

Gorski, J., Pfeuffer, F. & Klamroth, K. (2007), 'Biconvex sets and optimization with biconvex functions: a survey and extensions', *Mathematical Methods of Operations Research* **66**(3), 373–407.

Witten, D. M. & Tibshirani, R. (2011), 'Penalized classification using Fisher's linear discriminant', *Journal of the Royal Statistical Society Ser. B* **73**(5), 753–772.