Regularization oriented towards goal.

In many statistical applications, the ultimate goal, in inference about a covariance matrix, has to do with optimization of a quadratic form involving that unknown matrix. Denote the unknown matrix by $\Sigma = \Sigma_{m\times m}$, where $\Sigma = EZZ'$, $Z = (Z_1, ..., Z_m)$, $Z \sim F$.

For example principal component:

$$argmax_\beta \beta'\Sigma\beta \quad s.t. \; ||\beta|| = 1. \tag{1}$$

In case there is an independent sample $Z^i = (Z^i_1, ..., Z^i_m)$, $Z^i \sim F$, $i = 1, ..., n$, we define the obvious estimator S for Σ, $S = \frac{1}{n}\sum Z^i(Z^i)'$.

Note, when $m >> n$, there is no hope to find (even nearly) the maximizer of (1), but our goal is still to find $\hat{\beta}$ with high values of $\hat{\beta}'\Sigma\hat{\beta}$.

When replacing the original problem (1) by the following:

$$argmax_\beta \beta'S\beta \quad s.t. \; ||\beta|| = 1, \tag{2}$$

further regularizations are needed in order to obtain 'reliable' solutions. By reliable we mean that the empirical behavior of a solution $\hat{\beta}$ resembles the actual one, i.e. $\hat{\beta}'S\hat{\beta}$ is close to $\hat{\beta}'\Sigma\hat{\beta}$.

One popular method is the Lasso i.e., imposing a further constraint that $\sum |\beta_j| < C_1$. Here C_1 is a tuning parameter. The work in Greenshtein and Ritov (2005) and Greenshtein (2006), suggest that the value of C_1 is of the order of $\sqrt{n/\log(n)}$. But, in practice the exact value should be determined, e.g., through a test-set/ cross validation.

Thus, in Lasso we obtain a solution $\hat{\beta}$ to (2), but under the additional l_1 constraint.

I would like to suggest here an additional regularization constraint. Let $V(\beta) = Var(\beta'ZZ'\beta)$. Let $\hat{V}(\beta)$, be the obvious empirical estimator of $V(\beta)$. I suggest to add a constraint:

$$\hat{V}(\beta) < C_2,$$

where C_2 is again a tuning parameter that should be determined by a test-set/cross validation.

I believe that the solution $\hat{\beta}$ under both the l_1 and the additional constraint (tuned appropriately) is better (i.e., typically with higher values of $\hat{\beta}'\Sigma\hat{\beta}$), than the solution obtained only under a Lasso constraint. This should be true in a meaningful and large collection of setups.