SPARROW Equivalence

Christian Steffens

December 24, 2019

The main result in [SPP18] is that the $\ell_{2,1}$-mixed-norm regularized MMV problem

$$\boldsymbol{X}^\star = \arg\min_{\boldsymbol{X}} \frac{1}{2}\left\|\boldsymbol{A}\boldsymbol{X}-\boldsymbol{Y}\right\|_F^2 + \lambda\sqrt{N}\left\|\boldsymbol{X}\right\|_{2,1} \tag{1}$$

with $K \times N$ matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K]^\mathsf{T}$, is equivalent to the convex problem

$$\boldsymbol{S}^\star = \arg\min_{\boldsymbol{S}\in\mathbb{D}_+} \mathrm{Tr}\big((\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}^\mathsf{H} + \lambda\boldsymbol{I})^{-1}\hat{\boldsymbol{R}}\big) + \mathrm{Tr}\big(\boldsymbol{S}\big), \tag{2}$$

where $\boldsymbol{S} = \mathrm{diag}\left([s_1, \ldots, s_K]\right) \succcurlyeq \boldsymbol{0}$, in the sense that $\boldsymbol{X}^\star$ can be recovered by

$$\boldsymbol{X}^\star = \boldsymbol{S}^\star\boldsymbol{A}^\mathsf{H}(\boldsymbol{A}\boldsymbol{S}^\star\boldsymbol{A}^\mathsf{H} + \lambda\boldsymbol{I})^{-1}\boldsymbol{Y}. \tag{3}$$

A detailed derivation of the equivalence is given in the following section.

1 Proof of Equivalence

A key component in establishing the equivalence in equations (1)-(3) is the observation that the ℓ_2-norm of a vector $\boldsymbol{x}_k$, as used in problem (1), can be rewritten by means of factorization (cmp. [RFP10, Sec. 5.3]) as

$$\|\boldsymbol{x}_k\|_2 = \min_{\sigma_k, \boldsymbol{g}_k} \frac{1}{2}(\|\sigma_k\|_2^2 + \|\boldsymbol{g}_k\|_2^2) \quad \text{s.t. } \sigma_k\boldsymbol{g}_k = \boldsymbol{x}_k \tag{4}$$

where σ_k is a scalar and $\boldsymbol{g}_k$ is a vector of dimension $N \times 1$, similar to $\boldsymbol{x}_k$. We can extend this concept to the $\ell_{2,1}$-norm of the source signal matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K]^\mathsf{T}$ by

$$\sum_{n=1}^{N} \|\boldsymbol{x}_k\|_2 = \min_{\boldsymbol{\Sigma},\boldsymbol{G}} \frac{1}{2}(\|\boldsymbol{\Sigma}\|_F^2 + \|\boldsymbol{G}\|_F^2) \quad \text{s.t.} \boldsymbol{X} = \boldsymbol{\Sigma}\boldsymbol{G} \tag{5}$$

where $\Sigma = \mathrm{diag}\left(\sigma_1, \ldots, \sigma_k\right)$ is an $K \times K$ complex diagonal matrix and $\boldsymbol{G} = [\boldsymbol{g}_1, \ldots, \boldsymbol{g}_k]^\mathsf{T}$ is an $K \times N$ complex matrix with rows $\boldsymbol{g}_k$, for $k = 1, \ldots, K$. Inserting equation (5) into $\ell_{2,1}$-mixed-norm problem in (1) we formulate the bilinear optimization problem

$$\boldsymbol{\Sigma}^\star, \boldsymbol{G}^\star = \arg\min_{\boldsymbol{\Sigma},\boldsymbol{G}} \|\boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{G} - \boldsymbol{Y}\|_F^2 + \lambda\sqrt{N}(\|\boldsymbol{\Sigma}\|_F^2 + \|\boldsymbol{G}\|_F^2). \tag{6}$$

For fixed matrix $\boldsymbol{\Sigma}$ the problem in (6) has the closed form solution

$$\begin{aligned}\boldsymbol{G}^\star =& \arg\min_{\boldsymbol{G}} \|\boldsymbol{A\Sigma G}-\boldsymbol{Y}\|_F^2+\sqrt{N}\lambda(\|\boldsymbol{\Sigma}\|_F^2+\|\boldsymbol{G}\|_F^2)\\ =&(\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{A\Sigma}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{Y}\\ =&\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{A\Sigma\Sigma}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\boldsymbol{Y}\end{aligned}\tag{7}$$

where we made use of the Searle matrix inversion identity in the last equation. For ease of notation let us define

$$\boldsymbol{Z}=\boldsymbol{A\Sigma\Sigma}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}.\tag{8}$$

Using (8) and (7) the regularization term $\|\boldsymbol{G}^\star\|_F^2$ in (6) can be written as

$$\begin{aligned}\|\boldsymbol{G}^\star\|_F^2 =&\mathrm{Tr}\left(\boldsymbol{G}^\star\boldsymbol{G}^{\star\mathsf{H}}\right)\\ =&N\mathrm{Tr}\left((\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\boldsymbol{Z}(\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\hat{\boldsymbol{R}}\right)\end{aligned}\tag{9}$$

with $\boldsymbol{YY}^{\mathsf{H}}=N\hat{\boldsymbol{R}}$. Further using (8) and (7), the data fitting term in (6) yields

$$\begin{aligned}\|\boldsymbol{A\Sigma G}^\star-\boldsymbol{Y}\|_F^2 =&\left\|\boldsymbol{Z}(\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\boldsymbol{Y}-\boldsymbol{Y}\right\|_F^2\\ =&\left\|\lambda\sqrt{N}(\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\boldsymbol{Y}\right\|_F^2\\ =&\lambda^2N^2\mathrm{Tr}\left((\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-2}\hat{\boldsymbol{R}}\right).\end{aligned}\tag{10}$$

Combining (9) and (10) as

$$\begin{aligned}&\|\boldsymbol{A\Sigma G}^\star-\boldsymbol{Y}\|_F^2+\lambda\sqrt{N}\|\boldsymbol{G}^\star\|_F^2\\ =&\mathrm{Tr}(\lambda^2N^2(\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-2}\hat{\boldsymbol{R}}+\lambda N\sqrt{N}(\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\boldsymbol{Z}(\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\hat{\boldsymbol{R}})\\ =&\lambda N\sqrt{N}\ \mathrm{Tr}((\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}(\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})(\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\hat{\boldsymbol{R}})\\ =&\lambda N\sqrt{N}\ \mathrm{Tr}((\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\hat{\boldsymbol{R}})\end{aligned}\tag{11}$$

we can rewrite problem (6) as

$$\begin{aligned}\boldsymbol{\Sigma}^\star &= \arg\min_{\boldsymbol{\Sigma}}\ \lambda N\sqrt{N}\ \mathrm{Tr}((\boldsymbol{Z}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\hat{\boldsymbol{R}})+\lambda\sqrt{N}\mathrm{Tr}(\boldsymbol{\Sigma\Sigma}^{\mathsf{H}})\\ &= \arg\min_{\boldsymbol{\Sigma}}\ \lambda N\sqrt{N}\ \mathrm{Tr}((\boldsymbol{A\Sigma\Sigma}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\hat{\boldsymbol{R}})+\lambda\sqrt{N}\mathrm{Tr}(\boldsymbol{\Sigma\Sigma}^{\mathsf{H}}).\end{aligned}\tag{12}$$

Upon defining the nonnegative diagonal matrix $\boldsymbol{S}=\boldsymbol{\Sigma\Sigma}^{\mathsf{H}}/\sqrt{N}$ we arrive at the convex problem

$$\boldsymbol{S}^\star=\arg\min_{\boldsymbol{S}\succeq\boldsymbol{0}}\ \mathrm{Tr}((\boldsymbol{ASA}^{\mathsf{H}}+\lambda\boldsymbol{I})^{-1}\hat{\boldsymbol{R}})+\mathrm{Tr}(\boldsymbol{S}),\tag{13}$$

as given in (2). Make further use of the factorization in (5) to derive

$$\begin{aligned}\boldsymbol{X}^\star =&\boldsymbol{\Sigma}^\star\boldsymbol{G}^\star\\ =&\boldsymbol{\Sigma}^\star\boldsymbol{\Sigma}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{A\Sigma\Sigma}^{\mathsf{H}}\boldsymbol{A}^{\mathsf{H}}+\lambda\sqrt{N}\boldsymbol{I})^{-1}\boldsymbol{Y}\\ =&\boldsymbol{S}^\star\boldsymbol{A}^{\mathsf{H}}(\boldsymbol{AS}^\star\boldsymbol{A}^{\mathsf{H}}+\lambda\boldsymbol{I})^{-1}\boldsymbol{Y}\end{aligned}\tag{14}$$

which corresponds to the relation (3).

References

[RFP10] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[SPP18] C. Steffens, M. Pesavento, and M. E. Pfetsch. A compact formulation for the $\ell_{2,1}$ mixed-norm minimization problem. *IEEE Transactions on Signal Processing*, 66(6):1483–1497, March 2018.