Supplementary Material:
A Unified Optimization Framework for Low-Rank Inducing Penalties

Marcus Valtonen Örnhag[1] Carl Olsson[1,2]

[1]Centre for Mathematical Sciences
Lund University

[2]Department of Electrical Engineering
Chalmers University of Technology

{marcus.valtonen_ornhag, carl.olsson}@math.lth.se

1. Von Neumann's trace theorem

We use von Neumann's trace theorem repeatedly in the main paper, hence we state it here for completeness, using the inner products $\langle X, Y\rangle = \mathrm{tr}(\overline{X^T Y})$, and $\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}^T\mathbf{y}$.

Theorem 2 (Von Neumann [22]). *Let $X,\, Y \in \mathbb{C}^{n\times n}$ and $\boldsymbol{\sigma}(X)$ be the singular value vector of X. Then*

$$\langle X, Y\rangle \leq \langle \boldsymbol{\sigma}(X), \boldsymbol{\sigma}(Y)\rangle,$$

with equality if and only if X and Y are simultaneously unitarily diagonalizable.

Consider maximization over Z in (19) and note that

$$-\left\|X - Z\right\|_F^2 = -\left\|X\right\|_F^2 - \left\|Z\right\|_F^2 + 2\langle X, Z\rangle, \tag{40}$$

and by Theorem 2, $\langle X, Z\rangle \leq \langle \boldsymbol{\sigma}(X), \boldsymbol{\sigma}(Z)\rangle$, with equality if X and Z are simultaneously unitarily diagonalizable. Note that the Frobenius norm is unitarily invariant, with $\|X\|_F^2 = \sum_i \sigma_i(X)^2$. Therefore

$$-\left\|X - Z\right\|_F^2 \leq -\sum_i \left(\sigma_i(X) - \sigma_i(Z)\right)^2, \tag{41}$$

with equality if X and Z are simultaneously unitarily diagonalizable, i.e. $X = UD_{\boldsymbol{\sigma}(X)}V^T$ and $Z = UD_{\boldsymbol{\sigma}(Z)}V^T$, where $D_{\mathbf{x}}$ is a diagonal matrix with $\mathbf{x}$ on the main diagonal.

The remaining terms of (19) only depend on the singular values of Z and therefore the maximum occurs when we select Z so that we have equality in (41). This establishes the equality between (19) and (20) of the main paper.

2. The Fenchel Conjugate

In this section, we compute the Fenchel conjugate of (12), which is necessary in order to find the convex envelope. Let $\langle X, Y\rangle = \mathrm{tr}(X^T Y)$, and note that we can write

$$\langle Y, X\rangle - \left\|X - X_0\right\|_F = \left\|Z\right\|_F^2 - \left\|X_0\right\|_F^2 - \left\|Z - X\right\|_F^2. \tag{42}$$

where $Z = \frac{1}{2}Y + X_0$. By definition, the Fenchel conjugate of (12) is given by

$$\begin{aligned} f_h^*(Y) &= \sup_X \langle Y, X\rangle - f_h(X) \\ &= \sup_X \|Z\|_F^2 - \|X_0\|_F^2 - \|X - Z\|_F^2 - h(\boldsymbol{\sigma}(X)), \end{aligned} \tag{43}$$

where we use (42) in the last step. Note that the function h, as well as the Frobenius norm, is unitarily invariant. Furthermore, $\|X - Z\|_F^2 = \|X\|_F^2 + \|Z\|_F^2 - 2\langle X, Z\rangle$, and $\langle X, Z\rangle \leq \langle \boldsymbol{\sigma}(X), \boldsymbol{\sigma}(Z)\rangle$ by von Neumann's trace inequality, with equality if X and Z are simultaneously unitarily diagonalizable. This reduces the problem to optimizing over the singular values alone, which, after some manipulation, can be written as

$$\begin{aligned} f_h^*(Y) = \max_{\boldsymbol{\sigma}(X)} &-\|X_0\|_F^2 \\ &- \sum_{i=1}^k \left(\sigma_i^2(X) - 2[\sigma_i(Z) - a_i]\sigma_i(X) + b_i\right), \end{aligned} \tag{44}$$

where $\mathrm{rank}(X) = k$. Considering each singular value separately leads to a program on the form

$$\min_{x_i} x_i^2 - 2[\sigma_i(Z) - a_i]x_i + b_i, \tag{45}$$

subject to $\sigma_{i+1}(X) \leq x_i \leq \sigma_{i-1}(X)$. The sequence of unconstrained minimizers is given by $x_i = \sigma_i(Z) - a_i$. If there exists $x_i < 0$, then this is not the solution to the constrained problem. Nevertheless, the sequence is non-increasing, hence there is an index p, such that $x_p \geq 0$ and $x_{p+1} < 0$[4].

Note that

$$\sum_{i=1}^k x_i^2 - 2s_i x_i = \|\mathbf{x}\|^2 - 2\langle \mathbf{x}, \mathbf{s}\rangle, \tag{46}$$

[4]We allow the case $p = 0$, in which case the zero vector is optimal.

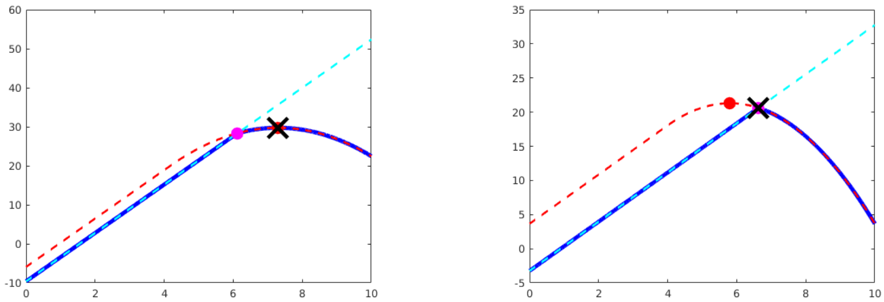

Figure 5. Two different cases for values of a_i, b_i and $\sigma_i(X)$ of (21).

hence we can consider optimizing $\|\mathbf{x}-\mathbf{s}\|^2 = \|\mathbf{x}\|^2 - 2\langle \mathbf{x}, \mathbf{s}\rangle + \|\mathbf{s}\|^2$ subject to $x_1 \geq x_2 \geq \cdots \geq x_k \geq 0$. Furthermore, $s_1 \geq s_2 \geq \cdots \geq s_k$.

Assume that minimum is obtained at $\mathbf{x}^*$ and fix x_p^*. Since $s_j < 0$ for all $j > p$, we must have $x_j^* = 0$ for $j > p$. It is now clear that, $x_j^* = s_j$ otherwise, hence $x_j^* = \max\{s_j, 0\} = [s_j]_+$. Inserting into (44) gives

$$f_h^*(Y) = \max_k -\|X_0\|_F^2 - \sum_{i=1}^{k} \left(b_i - [\sigma_i(Z) - a_i]_+^2\right). \tag{47}$$

Since $[s_i]_+ = [\sigma_i(Z) - a_i]_+$ is non-increasing, and b_i is non-decreasing, the maximizing k is obtained when

$$[\sigma_k(Z) - a_k]_+^2 \geq b_k \text{ and } b_{k+1} \geq [\sigma_{k+1}(Z) - a_{k+1}]_+^2. \tag{48}$$

For the maximizing $k = k^*$, we can write

$$\begin{aligned} &-\sum_{i=1}^{k^*} \left(b_i - [\sigma_i(Z) - a_i]_+^2\right) \\ &= \sum_{i=1}^{n} [\sigma_i(Z) - a_i]_+^2 - \sum_{i=1}^{n} \min\{b_i, [\sigma_i(Z) - a_i]_+^2\}. \end{aligned} \tag{49}$$

From this observation, we get

$$\begin{aligned} f_h^*(Y) = &\sum_{i=1}^{n} \left[\sigma_i(\frac{1}{2}Y + X_0) - a_i\right]_+^2 - \|X_0\|_F^2 \\ &- \sum_{i=1}^{n} \min\left(b_i, \left[\sigma_i(\frac{1}{2}Y + X_0) - a_i\right]_+^2\right). \end{aligned} \tag{50}$$

3. The Convex Envelope

Applying the definition of the bi-conjugate $f_h^{**}(X) = \sup_Y \langle Y, X\rangle - f_h^*(Y)$ to (50), and introduce the change of variables $Z = \frac{1}{2}Y + X_0$ we get

$$\begin{aligned} f_h^{**}(X) = &\max_Z 2\langle X, Z - X_0\rangle - \sum_{i=1}^{n} [\sigma_i(Z) - a_i]_+^2 \\ &+ \|X_0\|_F^2 + \sum_{i=1}^{n} \min\left(b_i, [\sigma_i(Z) - a_i]_+^2\right). \end{aligned} \tag{51}$$

By expanding squares and simplifying, $2\langle X, Z - X_0\rangle + \|X_0\|_F^2 = \|X - X_0\|_F^2 - \|X - Z\|_F^2 + \|Z\|_F^2$, which yields (19).

4. Obtaining the Maximizing Sequences

In this section we give the proof for the convergence of Algorithm 1, and how to modify it to cope with the corresponding problem for the proximal operator.

4.1. Proof of Theorem 1

Proof of Theorem 1. First, we will show that each step in the algorithm returns a solution to a constrained subproblem P_i, corresponding to a (partial) set of desired constraints $\mathcal{Z}_i$.

Let P_0 denote the unconstrained problem with solution $\int \in \mathbb{R}_+^n$. Denote the first interval generated in Algorithm 1 by $\iota_1 = \{m_1, \ldots, n_1\}$, and consider optimizing the first subproblem P_1

$$\max_{z_{m_1} \geq \cdots \geq z_{n_1}} c(\mathbf{z}), \tag{52}$$

where $\mathcal{Z}_1 = \{\mathbf{z} \in \mathcal{Z}_0 \mid z_{m_1} \geq \cdots \geq z_{n_1}\}$. By Lemma 1 the solution vector is constant over the subinterval $z_i = s$ for $i \in \iota_1$, which is returned by the algorithm. The next steps generates a solution to subproblem of the form

$$\max_{\substack{z_{m_1} \geq \cdots \geq z_{n_1} \\ \vdots \\ z_{m_k} \geq \cdots \geq z_{n_k}}} c(\mathbf{z}), \tag{53}$$

corresponding to subproblem P_k. I the solution to subproblem P_k is in $\mathcal{Z}$, then it is a solution to the problem, otherwise

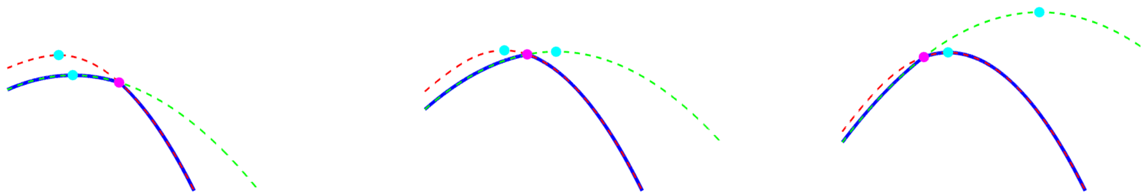

Figure 6. Illustration of the three different cases for the proximal operator.

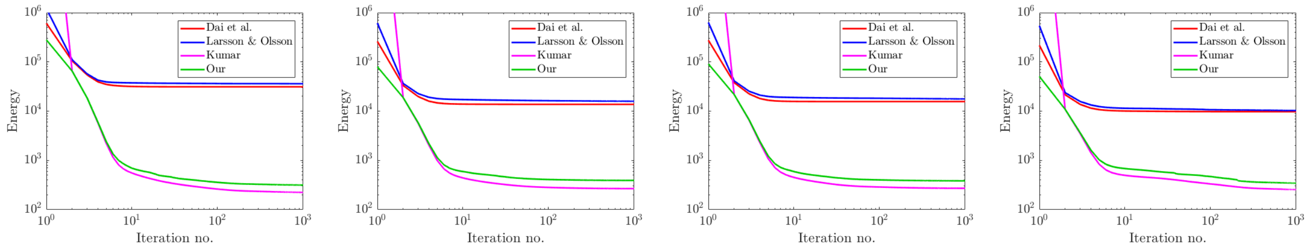

Figure 7. Convergence for the different methods compared in Section 6.3. NB: The energies are different, and have been been averaged over 35 different values of μ (the same values as in Figure 4).

one must add more constraints. We solve problems on the form

$$\max_{\mathbf{z}\in\mathcal{Z}_0} c(\mathbf{z}) \geq \max_{\mathbf{z}\in\mathcal{Z}_1} c(\mathbf{z}) \geq \cdots \geq \max_{\mathbf{z}\in\mathcal{Z}_\ell} c(\mathbf{z}) = \max_{\mathbf{z}\in\mathcal{Z}} c(\mathbf{z}), \tag{54}$$

where the last step yields a solution fulfilling the desired constraints. Furthermore $\mathcal{Z}_0 \supset \mathcal{Z}_1 \supset \cdots \supset \mathcal{Z}_\ell \supset \mathcal{Z}$, where $\mathcal{Z} = \{\mathbf{z} \mid z_1 \geq \cdots \geq z_n \geq 0\}$. Finally, it is easy to see that the algorithm terminates, since there are only finitely many possible subintervals. □

4.2. Modifying Algorithm 1

Following the approach used in [20], consider the program

$$\begin{aligned} \max_{s} \quad & \min\{b_i, [s-a_i]_+^2\} - \frac{\rho+1}{\rho}(s-\sigma_i(Y))^2 \\ & + s^2 - [s-a_i]_+^2, \\ \text{s.t.} \quad & \sigma_{i+1}(Z) \leq s \leq \sigma_{i-1}(Z). \end{aligned} \tag{55}$$

Note that the objective function is the pointwise minimum of

$$\begin{aligned} f_1(s) &= b_i - \frac{\rho+1}{\rho}(s-\sigma_i(Y))^2 + s^2 - [s-a_i]_+^2, \\ f_2(s) &= s^2 - \frac{\rho+1}{\rho}(s-\sigma_i(Y))^2, \end{aligned} \tag{56}$$

both of which are concave, since $\frac{\rho+1}{\rho} > 1$. For f_1 the maximum is obtained in $s = \frac{a_i\rho}{\rho+1} + \sigma_i(Y)$, if $s \geq a_i$ otherwise when $s = (\rho+1)\sigma_i(Y)$. The minimum of f_2 is obtained when $s = (\rho+1)\sigma_i(Y)$.

There are three possible cases, also shown in Figure 6.

1. The maximum occurs when $s > a_i + \sqrt{b_i}$, hence $f_1(s) < f_2(s)$, hence $s = \frac{a_i\rho}{\rho+1} + \sigma_i(Y)$.
2. The maximum occurs when $s < a_i + \sqrt{b_i}$, where $f_1(s) > f_2(s)$, hence $s = (\rho+1)\sigma_i(Y)$.
3. When $s = a_i + \sqrt{b_i}$, which is valid elsewhere.

in summary

$$s_i = \begin{cases} \dfrac{a_i\rho}{\rho+1} + \sigma_i(Y), & \dfrac{a_i}{\rho+1} + \sqrt{b_i} < \sigma_i(Y), \\ a_i + \sqrt{b_i}, & \dfrac{a_i+\sqrt{b_i}}{1+\rho} \leq \sigma_i(Y) \leq \dfrac{a_i}{\rho+1} + \sqrt{b_i}, \\ (1+\rho)\sigma_i(Y), & \sigma_i(Y) < \dfrac{a_i+\sqrt{b_i}}{1+\rho}, \end{cases} \tag{57}$$

By replacing the sequence of unconstrained minimizers $\{s_i\}$ defined by (57), with the corresponding sequence in Section 4 (of the main paper), and changing the objective function of Algorithm 1, to the one in (55), the maximizing singular value vector fo the proximal operator is obtained.

5. Convergence: Motion Capture

In this section we compare the convergence of the different regularizers used in Section 6.3, see Figure 7. Note that the energies are different, and one can only compare the number of steps needed until convergence. For this particular choice of a_i and b_i the $\mathcal{R}_h$ regularizer behaves much like WNNM used in [17].

References

[1] Roland Angst, Christopher Zach, and Marc Pollefeys. The generalized trace-norm and its application to structure-from-motion problems. In *International Conference on Computer Vision*, 2011. 1

[2] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, May 2007. 1

[3] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, 2011. 3, 4, 5

[4] C. Bregler, A. Hertzmann, and H. Biermann. Recovering non-rigid 3d shape from image streams. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2000. 1

[5] R. Cabral, F. De la Torre, J. P. Costeira, and A. Bernardino. Unifying nuclear norm and bilinear factorization approaches for low-rank matrix decomposition. In *International Conference on Computer Vision (ICCV)*, 2013. 5

[6] Emmanuel J Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011. 1

[7] Emmanuel J. Candès, Xiaodong Li, Yi Ma, and John Wright. Robust principal component analysis? *J. ACM*, 58(3):11:1–11:37, 2011. 5

[8] Emmanuel J Candes, Michael B Wakin, and Stephen P Boyd. Enhancing sparsity by reweighted l^1 minimization. *Journal of Fourier analysis and applications*, 14(5-6):877–905, 2008. 5

[9] Lu Canyi, Jinhui Tang, Shuicheng Yan, and Zhouchen Lin. Nonconvex nonsmooth low-rank minimization via iteratively reweighted nuclear norm. *IEEE Transactions on Image Processing*, 25, 10 2015. 1, 5

[10] Marcus Carlsson. On convexification/optimization of functionals including an l2-misfit term. *arXiv preprint arXiv:1609.09378*, 2016. 3

[11] Marcus Carlsson, Daniele Gerosa, and Carl Olsson. An unbiased approach to compressed sensing. *arXiv preprint*, arXiv:1806.05283, 2018. 3

[12] Yuchao Dai, Hongdong Li, and Mingyi He. A simple prior-free method for non-rigid structure-from-motion factorization. *International Journal of Computer Vision*, 107(2):101–122, 2014. 3, 6, 7

[13] Ravi Garg, Anastasios Roussos, and Lourdes Agapito. A variational approach to video registration with subspace constraints. *International Journal of Computer Vision*, 104(3):286–314, 2013. 1

[14] N. Gillis and F. Glinuer. Low-rank matrix approximation with weights or missing data is np-hard. *SIAM Journal on Matrix Analysis and Applications*, 32(4), 2011. 1

[15] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International Journal of Computer Vision*, 121, 07 2016. 2, 5, 6

[16] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He. Fast and accurate matrix completion via truncated nuclear norm regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2117–2130, 2013. 1

[17] Suryansh Kumar. Non-rigid structure from motion: Prior-free factorization method revisited. In *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2020. 5, 6, 7, 13

[18] Suryansh Kumar, Yuchao Dai, and Hongdong Li. Superpixel soup: Monocular dense 3d reconstruction of a complex dynamic scene. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 2019. 3

[19] Viktor Larsson, Erik Bylow, Carl Olsson, and Fredrik Kahl. Rank minimization with structured data patterns. In *European Conference on Computer Vision*, 2014. 2, 6

[20] Viktor Larsson and Carl Olsson. Convex low rank approximation. *International Journal of Computer Vision*, 120(2):194–214, 2016. 1, 2, 3, 4, 5, 7, 13

[21] Viktor Larsson and Carl Olsson. Compact matrix factorization with dependent subspaces. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4361–4370, 07 2017. 6

[22] L. Mirsky. A trace inequality of john von neumann. *Monatshefte fr Mathematik*, 79:303–306, 1975. 2, 11

[23] Karthik Mohan and Maryam Fazel. Iterative reweighted least squares for matrix rank minimization. In *Annual Allerton Conference on Communication, Control, and Computing*, pages 653–661, 2010. 1

[24] R. A. Newcombe, D. Fox, and S. M. Seitz. Dynamicfusion: Reconstruction and tracking of non-rigid scenes in real-time. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 343–352, June 2015. 3

[25] Feiping Nie, Hua Wang, Xiao Cai, Heng Huang, and Chris H. Q. Ding. Robust matrix completion via joint schatten p-norm and lp-norm minimization. In *ICDM*, pages 566–574, 2012. 5

[26] T. H. Oh, Y. W. Tai, J. C. Bazin, H. Kim, and I. S. Kweon. Partial sum minimization of singular values in robust pca: Algorithm and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(4):744–758, 2016. 1

[27] Carl Olsson, Marcus Carlsson, Fredrik Andersson, and Viktor Larsson. Non-convex rank/sparsity regularization and local minima. *Proceedings of the International Conference on Computer Vision*, 2017. 1, 3, 6

[28] Carl Olsson, Marcus Carlsson, and Daniele Gerosa. Bias reduction in compressed sensing. *arXiv preprint*, arxiv:1812.11329, 2018. 2

[29] S. Oymak, A. Jalali, M. Fazel, Y. C. Eldar, and B. Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. *IEEE Transactions on Information Theory*, 61(5):2886–2908, 2015. 1

[30] Benjamin Recht, Maryam Fazel, and Pablo A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.*, 52(3):471–501, Aug. 2010. 1

[31] C. Russell, R. Yu, and L. Agapito. Video-popup: Monocular 3d reconstruction of dynamic scenes. In *European Conference on Computer Vision (ECCV)*, 2014. 3

[32] F. Shang, J. Cheng, Y. Liu, Z. Luo, and Z. Lin. Bilinear factor matrix norm minimization for robust pca: Algorithms and applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2066–2080, Sep. 2018. 1, 5

[33] Jonathan Taylor, Allan Jepson, and Kiriakos Kutulakos. Non-rigid structure from locally-rigid motion. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2761–2768, 2010. 3

[34] Kim-Chuan Toh and Sangwoon Yun. An accelerated proximal gradient algorithm for nuclear norm regularized least squares problems. *Pacific Journal of Optimization*, 6, 09 2010. 5

[35] J. Valmadre, S. Sridharan, S. Denman, C. Fookes, and S. Lucey. Closed-form solutions for low-rank non-rigid reconstruction. In *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–6, Nov 2015. 6

[36] Jun Xu, Lei Zhang, David Zhang, and Xiangchu Feng. Multi-channel weighted nuclear norm minimization for real color image denoising. *International Conference on Computer Vision (ICCV)*, 2017. 2

[37] Noam Yair and Tomer Michaeli. Multi-scale weighted nuclear norm image restoration. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2