Supplementary Material:
Bilinear Parameterization for Non-Separable Singular Value Penalties

Marcus Valtonen Örnhag[1] José Pedro Iglesias[2] Carl Olsson[1,2]

[1]Centre for Mathematical Sciences
Lund University

[2]Department of Electrical Engineering
Chalmers University of Technology

1. Proof of Lemma 1

We present the omitted proof from the main paper, but first we establish the following property.

Lemma 2. *The sequence $(s_i - \gamma_i)_{i=1}^n$ is non-decreasing.*

Proof. We choose k such that $\gamma_i \geq \sqrt{b_i}$ for $i \leq k$ and $\gamma_i < \sqrt{b_i}$ for $i > k$. Then, for $i < k$

$$\begin{aligned} s_i - \gamma_i &= a_i + \max(\gamma_i, \sqrt{b_i}) - \gamma_i \\ &= a_i \leq a_{i+1} = s_{i+1} - x_{i+1}, \end{aligned} \tag{42}$$

and, for $i = k$

$$\begin{aligned} s_k - x_k &= a_k \leq a_{k+1} \leq a_{k+1} + \sqrt{b_{k+1}} - x_{k+1} \\ &= s_{k+1} - x_{k+1}. \end{aligned} \tag{43}$$

For $i > k$ we have

$$s_i - \gamma_i = a_i + \sqrt{b_i} - \gamma_i, \tag{44}$$

which is non-decreasing since both of $(a_i + \sqrt{b_i})_{i=1}^n$ and $(-\gamma_i)_{i=1}^n$ are. □

Now, we turn our attention to the original statement.

Proof of Lemma 1. First we note that since the elements of $\mathbf{v}$ are positive the elements of $\mathbf{z}^*$ that can take multiple values should be taken as large as possible. As stated above this means that (25) holds for all elements of $\mathbf{z}^*$. If $z_i^* = z_{i+1}^*$ then

$$z_i^* - \gamma_i = z_{i+1}^* - \gamma_{i+1} + \underbrace{\gamma_{i+1} - \gamma_i}_{\leq 0} \leq z_{i+1}^* - \gamma_{i+1} \tag{45}$$

If $z_i^* > z_{i+1}^*$ then according to (25) we have $z_i^* \leq s_i$ and $z_{i+1}^* \geq s_{i+1}$ which, together with Lemma 2, gives

$$z_i^* - \gamma_i \leq s_i - \gamma_i \leq s_{i+1} - \gamma_{i+1} \leq z_{i+1}^* - \gamma_{i+1}. \tag{46}$$

□

1.1. Proof of the claim in Section 2.3

We formulate the claim as a lemma.

Lemma 3. *Let $2\mathbf{z} \in \partial g(\boldsymbol{\eta})$. Then z_i can only take one value when $\eta_i \neq 0$, for all $i = 1, \ldots, n$.*

Proof. By Danskin's Theorem (see *e.g.* [3]) it follows that $\partial g(\boldsymbol{\eta}) = 2 \arg\max_{\mathbf{z}} \ell(\boldsymbol{\eta}, \mathbf{z})$, where

$$\ell(\boldsymbol{\eta}, \mathbf{z}) := 2\langle \boldsymbol{\eta}, \mathbf{z} \rangle - \sum_{i=1}^n \left[\left[|z_{[i]}| - a_i \right]_+^2 - b_i \right]_+, \tag{47}$$

such that $g(\boldsymbol{\eta}) = \max_z \ell(\boldsymbol{\eta}, \mathbf{z})$. Partition $\boldsymbol{\eta} = (\boldsymbol{\eta}', \mathbf{0})$, where all elements of $\boldsymbol{\eta}' \in \mathbb{R}_k$, $k \leq n$, has a positive magnitude. It is clear that we may choose any value $|z_{[i]}| \leq a_i + \sqrt{b_i}$ for the last $n - k$ elements of the maximizing vector, as long as $(|z_{[i]}|)_{i=1}^n$ is non-increasing. Furthermore, the first k elements are obtained by

$$\ell(\boldsymbol{\eta}', \mathbf{z}') := 2\langle \boldsymbol{\eta}', \mathbf{z}' \rangle - \sum_{i=1}^k \left[\left[|z'_{[i]}| - a_i \right]_+^2 - b_i \right]_+, \tag{48}$$

where $\mathbf{z}' \in \mathbb{R}^k$ are the first k elements of $\mathbf{z}$. The objective in (48) is strictly concave in $\mathbf{z}'$, since all $\eta_i' > 0$, and, consequently, has a unique maximizer. In turn, the elements of the maximizing vector $\mathbf{z}$, can only take a single value, when $\eta \neq 0$.

□

2. Detailed Algorithm

In this section we present a more detailed description of the algorithm. First we approximate the regularization term, as described in Section 2.3, then apply the Ruhe Wedin 2 approximation (see [19] for details) and take one step of VarPro. Since VarPro uses Jacobians, w.r.t. B and C, we must linearize the contribution from the regularizer. As the regularization term becomes

$$\| \operatorname{diag}(w^{(t)}) B \|_F^2 + \| \operatorname{diag}(w^{(t)}) C \|_F^2, \tag{49}$$

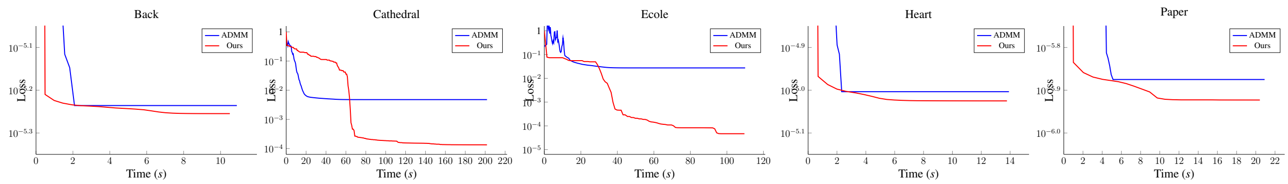

Figure 6: Converge plots for the remaining sequences.

where $\mathrm{diag}(w^{(t)})$ is a diagonal matrix with the weights $w_i^{(t)}$ in the diagonal, a natural linearization emerges by simply column stacking the variables B and C, respectively.

With $\mathbf{b} = \mathrm{vec}(B)$ and $\mathbf{c} = \mathrm{vec}(C^T)$, we may write (49) as

$$\|J_B^{\text{reg}}\mathbf{b}\|^2 + \|J_C^{\text{reg}}\mathbf{c}\|^2. \tag{50}$$

Note that J_B^{reg} and J_C^{reg} are diagonal matrices.

Given a current iterate $(\mathbf{b}^{(t)}, \mathbf{c}^{(t)})$ we write the regularization term as $\|J_B^{\text{reg}}\delta\mathbf{b} + \mathbf{r}_B\|^2 + \|J_C^{\text{reg}}\delta\mathbf{c} + \mathbf{r}_C\|^2$, where $\mathbf{r}_B = J_B^{\text{reg}}\mathbf{b}^{(t)}$, $\mathbf{r}_C = J_C^{\text{reg}}\mathbf{c}^{(t)}$, $\mathbf{b} = \mathbf{b}^{(t)} + \delta\mathbf{b}$ and $\mathbf{c} = \mathbf{c}^{(t)} + \delta\mathbf{c}$.

For the data term, the residuals $\mathcal{A}BC^T - b$ around $(\mathbf{b}^{(t)}, \mathbf{c}^{(t)})$ can be linearized, resulting in

$$J_B^{\text{data}}\delta\mathbf{b} + J_C^{\text{data}}\delta\mathbf{c} + \mathbf{r}^{\text{data}}. \tag{51}$$

The full objective can therefore be written as

$$\|J_B\delta\mathbf{b} + J_C\delta\mathbf{c} + \mathbf{r}\|^2, \tag{52}$$

where

$$J_B = \begin{bmatrix} J_B^{\text{reg}} \\ 0 \\ J_B^{\text{data}} \end{bmatrix},\ J_C = \begin{bmatrix} 0 \\ J_C^{\text{reg}} \\ J_C^{\text{data}} \end{bmatrix},\ \mathbf{r} = \begin{bmatrix} \mathbf{r}_B \\ \mathbf{r}_C \\ \mathbf{r}^{\text{data}} \end{bmatrix}. \tag{53}$$

Lastly, an optional refactorizing of the current iterate using SVD can be performed, as discussed in the main paper. We summarize these steps in Algorithm 1.

Input: Robust penalty function f, linear operator $\mathcal{A}$ and regularization parameter μ, damping parameter λ.
Initialize B and C with random entries
while *not converged* **do**
  Compute weights $w^{(t)}$ from current iterate (B, C)
  Compute the vectorizations $\mathbf{b} = \mathrm{vec}(B)$, $\mathbf{c} = \mathrm{vec}(C^T)$
  Compute residuals $\mathbf{r}_B$ $\mathbf{r}_C$, and Jacobians J_B^{data} and J_B^{data} depending on $\mathcal{A}$
  Compute residual $\mathbf{r}^{\text{reg}}$, and Jacobians J_B^{reg} and J_C^{reg}
  Create full residual $\mathbf{r}$ and Jacobians J_B and J_C
  Compute $\tilde{J}^T\tilde{J} + \lambda I = J_B^T(I - J_C J_C^+)J_B + \lambda I$
  Compute $\mathbf{b}' = \mathbf{b} - (\tilde{J}^T\tilde{J} + \lambda I)^{-1}J_B r$ and reshape into matrix B'
  Compute C' by minimize the full objective (52) with fixed B'
  if $\mathcal{R}(B'C'^T) + \|\mathcal{A}(B'C'^T) - b\|^2 < \mathcal{R}(BC^T) + \|\mathcal{A}(BC^T) - b\|^2$ **then**
    $[U, \Sigma, V] = \mathrm{svd}(B'C'^T)$
    Update $B = U\sqrt{\Sigma}$ and $C = V\sqrt{\Sigma}$
    Decrease λ
  else
    Increase λ
  end
end

Algorithm 1: Outline of the bilinear method.

3. Parameters Used in Experiments

The values of the parameters used to define the sequences for each of the datasets are shown in Table 3. In all experiments we use $\delta = 10^{-6}$.

4. Convergence plots

We display the convergence plots that were omitted in the main paper, see Figure 6. As was noted in the main text, the difference in loss is substantially larger for rigid objects; however, even small differences in loss can have a significant impact on 3D reconstructions.

Table 3: Parameters used for defining the sequences $\{a_i\}$ and $\{b_i\}$ in the matrix recovery experiments.†For the Cathedral, École, and Door datasets with WNN regularization we set $a_i = 0$ for the first four singular values and a high value for the remaining ones (truncated nuclear norm).

	a_{NN}	a_{WNN}	a_{LI}	a_{SV}
Cathe.	7.5×10^{-3}	$0/10^\dagger$	-	-
École	2.5×10^{-3}	$0/10^\dagger$	-	-
Door	1.5×10^{-3}	$0/10^\dagger$	-	-
Back	5×10^{-4}	1×10^{-6}	5×10^{-6}	5×10^{-9}
Heart	4.5×10^{-4}	5×10^{-6}	5×10^{-6}	1×10^{-8}
Paper	2.6×10^{-4}	2.5×10^{-7}	1×10^{-6}	5×10^{-10}

References

[1] Francis R. Bach. Convex relaxations of structured matrix factorizations. *CoRR*, abs/1309.3117, 2013. 3

[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] D.P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999. 11

[4] Rajendra Bhatia and Tanvi Jain. On symplectic eigenvalues of positive definite matrices. *Journal of Mathematical Physics*, 56(11):112201, 2015. 6

[5] 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, Jan. 2011. 2

[6] 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

[7] A. M. Buchanan and A. W. Fitzgibbon. Damped newton algorithms for matrix factorization with missing data. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2005. 1

[8] 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. 1, 2, 3

[9] Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, 2009. 2

[10] Lu Canyi, Jinhui Tang, Shuicheng Yan, and Zhouchen Lin. Generalized nonconvex nonsmooth low-rank minimization. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014. 1, 2

[11] Marcus Carlsson. On convex envelopes and regularization of non-convex functionals without moving global minima. *Journal of Optimization Theory and Applications, to appear*, 2019. 2, 3, 4

[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. 1, 2, 8

[13] Anders Eriksson, Trung Thanh Pham, Tat-Jun Chin, and Ian Reid. The k-support norm and convex envelopes of cardinality and rank. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2

[14] Christian Grussler and Pontus Giselsson. Low-rank inducing norms with optimality interpretations. *CoRR*, abs/1612.03186, 2016. 2

[15] Christian Grussler, Anders Rantzer, and Pontus Giselsson. Low-rank optimization with convex constraints. *CoRR*, abs/1606.01793, 2016. 2

[16] Benjamin D. Haeffele and René Vidal. Structured low-rank matrix factorization: Global optimality, algorithms, and applications. *CoRR*, abs/1708.07850, 2017. 1, 3

[17] J. He, L. Balzano, and A. Szlam. Incremental gradient on the grassmannian for online foreground and background separation in subsampled video. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1568–1575, June 2012. 2

[18] Je Hyeong Hong and Andrew Fitzgibbon. Secrets of matrix factorization: Approximations, numerics, manifold optimization and random restarts. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. 2, 5

[19] Je Hyeong Hong, Christopher Zach, and Andrew Fitzgibbon. Revisiting the variable projection method for separable nonlinear least squares problems. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2, 5, 11

[20] Je Hyeong Hong, Christopher Zach, Andrew Fitzgibbon, and Roberto Cipolla. Projective bundle adjustment from arbitrary initialization using the variable projection method. In *European Conference on Computer Vision (ECCV)*, 2016. 2, 5

[21] 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, 2

[22] Je Hyeong Hong and Christopher Zach. pose: Pseudo object space error for initialization-free bundle adjustment. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 1, 2, 5, 7

[23] José Pedro Iglesias, Carl Olsson, and Marcus Valtonen Örnhag. Accurate optimization of weighted nuclear norm for non-rigid structure from motion. In *Proceedings of the European Computer Vision Conference (ECCV)*, 2020. 3, 5, 6, 7, 8

[24] Sebastian Hoppe Nesgaard Jensen, Alessio Del Bue, Mads Emil Brix Doest, and Henrik Aanæs. A benchmark and evaluation of non-rigid structure from motion, 2018. 8

[25] Suryansh Kumar. A simple prior-free method for non-rigid structure-from-motion factorization : Revisited. *CoRR*, abs/1902.10274, 2019. 1, 8

[26] Suryansh Kumar. Non-rigid structure from motion: Prior-free factorization method revisited. In *IEEE Winter Conference on Applications of Computer Vision, WACV 2020, Snowmass Village, CO, USA, March 1-5, 2020*, pages 51–60. IEEE, 2020. 1, 2

[27] Hanjiang Lai, Yan Pan, Canyi Lu, Yong Tang, and Shuicheng Yan. Efficient k-support matrix pursuit. In *European Conference on Computer Vision 2014*, volume 8690, 2014. 2

[28] Canyi Lu, Changbo Zhu, Chunyan Xu, Shuicheng Yan, and Zhouchen Lin. Generalized singular value thresholding. In *AAAI*, 2015. 1, 2

[29] Andrew M McDonald, Massimiliano Pontil, and Dimitris Stamos. Spectral k-support norm regularization. In *Advances in Neural Information Processing Systems*. 2014. 2

[30] 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, 2

[31] 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. 4

[32] C. Olsson, M. Carlsson, and E. Bylow. A non-convex relaxation for fixed-rank approximation. In *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 1809–1817, 2017. 4

[33] Carl Olsson and Olof Enqvist. Stable structure from motion for unordered image collections. In *Proceedings of the Scandinavian Conference on Image Analysis (SCIA)*, pages 524–535, 2011. 8

[34] Marcus Valtonen Örnhag, Carl Olsson, and Anders Heyden. Differentiable fixed-rank regularisationusing bilinear parameterisation. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2019. 3

[35] 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. 2, 4

[36] Chris Russell, Joao Fayad, and Lourdes Agapito. Energy based multiple model fitting for non-rigid structure from motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3009 – 3016, 07 2011. 8

[37] 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. 3

[38] Danail Stoyanov, George P. Mylonas, Fani Deligianni, Ara Darzi, and Guang Zhong Yang. Soft-tissue motion tracking and structure estimation for robotic assisted mis procedures. In James S. Duncan and Guido Gerig, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*, pages 139–146, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg. 8

[39] Carlo Tomasi and Takeo Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992. 1

[40] 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. 8

[41] Marcus Valtonen Örnhag and Carl Olsson. A unified optimization framework for low-rank inducing penalties. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 2, 4, 5, 6, 7

[42] Marcus Valtonen Örnhag, Carl Olsson, and Anders Heyden. Bilinear parameterization for differentiable rank-regularization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020. 3

[43] Aydin Varol, Mathieu Salzmann, Engin Tola, and Pascal Fua. Template-free monocular reconstruction of deformable surfaces. In *International Conference on Computer Vision (ICCV)*, pages 1811 – 1818, 11 2009. 8

[44] Naiyan Wang, Tiansheng Yao, Jingdong Wang, and Dit-Yan Yeung. A probabilistic approach to robust matrix factorization. In *European Conferencec on Computer Vision (ECCV)*, 2012. 1

[45] Chen Xu, Zhouchen Lin, and Hongbin Zha. A unified convex surrogate for the schatten-p norm. In *Proceedings of the Conference on Artificial Intelligence (AAAI)*, 2017. 3

[46] Noam Yair and Tomer Michaeli. Multi-scale weighted nuclear norm image restoration. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1