Rank constrained regression

Rajarshi Saha

August 9, 2024

Main result

Rank-constrained regression refers to the following least squares minimization problem, where the (matrix) regressor is constrained to be low-rank.

$$\min_{\mathrm{rank}(\mathbf{Z}) \leq k} \|\mathbf{XZ} - \mathbf{Y}\|_{\mathrm{F}}^2 . \tag{1}$$

A low-rank matrix admits an **LR** factorization, where **L** and **R** are tall and wide matrices respectively. Although problem (1) is non-convex, it can be solved to global optimality via two SVDs. This result likely first appears in [1]. What follows below is a re-derivation of the result with simplified notation, just in terms of singular values, and copied verbatim from [2, Lemma B.1].

Suppose $\mathbf{Y} \in \mathbb{R}^{m \times n}$ is given, and suppose $\mathbf{X} \in \mathbb{R}^{m \times d}$ is full rank, i.e., $\mathrm{rank}(\mathbf{X}) = \min\{m, d\}$, with SVD, $\mathbf{X} = \mathbf{U}\widetilde{\mathbf{\Sigma}}\mathbf{V}^\top$. Furthermore, let $\grave{\mathbf{U}}\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top$ denote the full SVD of $\mathbf{U}^\top\mathbf{Y}$ if $m \leq d$, or the full SVD of $\left(\mathbf{U}\mathbf{I}_d\right)^\top \mathbf{Y}$ if $m > d$. Then, for $k \leq m$, the solution of (1) is given by

$$\mathbf{Z}_* := \arg\min_{\mathrm{rank}(\mathbf{Z}) \leq k} \|\mathbf{XZ} - \mathbf{Y}\|_{\mathrm{F}}^2 = \begin{cases} \left(\mathbf{V}\mathbf{I}_m\mathbf{\Sigma}^{-1}\grave{\mathbf{U}}\mathbf{I}_k\right)\left(\mathbf{I}_k^\top\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top\right) & \text{if} \quad m \leq d, \\ \left(\mathbf{V}\mathbf{\Sigma}^{-1}\grave{\mathbf{U}}\mathbf{I}_k\right)\left(\mathbf{I}_k^\top\grave{\mathbf{\Sigma}}\mathbf{V}^\top\right) & \text{otherwise.} \end{cases} \tag{2}$$

Here, $\mathbf{\Sigma}$ is a diagonal matrix of the non-zero singular values of $\mathbf{X}$, defined as $\mathbf{\Sigma} := \widetilde{\mathbf{\Sigma}}\mathbf{I}_m \in \mathbb{R}^{m \times m}$ when $m \leq d$, and $\mathbf{\Sigma} := \mathbf{I}_d^\top\widetilde{\mathbf{\Sigma}} \in \mathbb{R}^{d \times d}$ when $d < m$. Additionally, the optimal value is

$$\begin{aligned} &\min_{\mathrm{rank}(\mathbf{Z}) \leq k} \|\mathbf{XZ} - \mathbf{Y}\|_{\mathrm{F}}^2 \\ &\quad = \|\mathbf{XZ}_* - \mathbf{Y}\|_{\mathrm{F}}^2 = \begin{cases} \sum_{i=k+1}^{m} \sigma_i^2\left(\mathbf{Y}\right), & \text{if} \quad m \leq d, \\ \sum_{i=k+1}^{m} \sigma_i^2\left(\left(\mathbf{U}\mathbf{I}_d\right)^\top\mathbf{Y}\right) + \left\|\left(\mathbf{U}\bar{\mathbf{I}}_{m-d}\right)^\top\mathbf{Y}\right\|_{\mathrm{F}}^2, & \text{otherwise.} \end{cases} \end{aligned} \tag{3}$$

Proof

Case $\mathbf{m} \leq \mathbf{d}$: Since the full SVD of $\mathbf{X} \in \mathbb{R}^{m \times d}$ is $\mathbf{X} = \mathbf{U}\widetilde{\mathbf{\Sigma}}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{m \times m}$, $\widetilde{\mathbf{\Sigma}} \in \mathbb{R}^{m \times d}$, and $\mathbf{V} \in \mathbb{R}^{d \times d}$, the last $(d-m)$ columns of $\widetilde{\mathbf{\Sigma}}$ will be zero, i.e., $\widetilde{\mathbf{\Sigma}}\bar{\mathbf{I}}_{m-d} = \mathbf{0}$. Let $\mathbf{Z}' := \mathbf{V}^\top\mathbf{Z} \in \mathbb{R}^{d \times n}$ be the transformed optimization variable. Since $\mathbf{V}$ is a unitary matrix, $\mathrm{rank}(\mathbf{Z}) \leq k$ if and only if $\mathrm{rank}(\mathbf{Z}') \leq k$. Splitting $\mathbf{Z}' \in \mathbb{R}^{d \times n}$ into $\mathbf{Z}'' := \mathbf{I}_m^\top\mathbf{Z}' \in \mathbb{R}^{m \times n}$ and $\overline{\mathbf{Z}}'' := \bar{\mathbf{I}}_{d-m}^\top\mathbf{Z}' \in \mathbb{R}^{(d-m) \times n}$, it can be seen that $\widetilde{\mathbf{\Sigma}}\mathbf{Z}' =$

$\widetilde{\mathbf{\Sigma}}\mathbf{I}_d\mathbf{Z}'' + \widetilde{\mathbf{\Sigma}}\overline{\mathbf{I}}_{m-d}\overline{\mathbf{Z}}'' = \mathbf{\Sigma}\mathbf{Z}''$, where $\mathbf{\Sigma} := \widetilde{\mathbf{\Sigma}}\mathbf{I}_m \in \mathbb{R}^{m\times m}$ is a diagonal matrix comprised of the non-zero singular values of $\mathbf{X}$. Then,

$$\min_{\text{rank}(\mathbf{Z})\le k} \|\mathbf{X}\mathbf{Z} - \mathbf{Y}\|_{\text{F}}^2 \equiv \min_{\text{rank}(\mathbf{Z}')\le k} \left\|\widetilde{\mathbf{\Sigma}}\mathbf{Z}' - \mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2 \equiv \min_{\text{rank}(\mathbf{Z}')\le k} \left\|\mathbf{\Sigma}\mathbf{Z}'' - \mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2. \quad (4)$$

Note that the objective function value is independent of $\overline{\mathbf{Z}}''$, as $\overline{\mathbf{Z}}''$ lies in the null space of $\mathbf{X}^\top$. Since $\mathbf{X}$ is full rank, $\mathbf{\Sigma}$ is invertible, and the minimization in (4) is equivalent to

$$\min_{\text{rank}(\widetilde{\mathbf{Z}})\le \text{k}} \left\|\widetilde{\mathbf{Z}} - \mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2 = \sum_{i=k+1}^{m} \sigma_i^2\left(\mathbf{Y}\right), \quad (5)$$

where $\widetilde{\mathbf{Z}} := \mathbf{\Sigma}\mathbf{Z}'' \in \mathbb{R}^{m\times n}$. The equality in (5) follows as a consequence of Eckart-Young-Mirsky theorem, which states that $\|\widetilde{\mathbf{Z}} - \mathbf{U}^\top\mathbf{Y}\|_{\text{F}}^2$ is minimized by taking the best rank-k approximation of $\mathbf{U}^\top\mathbf{Y}$, and the fact that the singular values of $\mathbf{U}^\top\mathbf{Y}$ and $\mathbf{Y}$ are the same as $\mathbf{U}$ is unitary. Moreover, as the SVD of $\mathbf{U}^\top\mathbf{Y}$ is $\grave{\mathbf{U}}\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top$,

$$\begin{aligned} \widetilde{\mathbf{Z}}_* &:= \arg\min_{\text{rank}(\widetilde{\mathbf{Z}})\le \text{k}} \left\|\widetilde{\mathbf{Z}} - \mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2 = \left(\grave{\mathbf{U}}\mathbf{I}_k\right)\left(\mathbf{I}_k^\top\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top\right) \\ \text{and,} \quad \mathbf{Z}''_* &:= \mathbf{\Sigma}^{-1}\widetilde{\mathbf{Z}}_* = \left(\mathbf{\Sigma}^{-1}\grave{\mathbf{U}}\mathbf{I}_k\right)\left(\mathbf{I}_k^\top\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top\right). \end{aligned} \quad (6)$$

In other words, if $\mathbf{Z}'_*$ denotes the solution of the middle optimization problem in (4), $\mathbf{I}_m^\top\mathbf{Z}_* = \mathbf{Z}''_*$. Furthermore, note that setting $\overline{\mathbf{I}}_{d-m}^\top\widetilde{\mathbf{Z}}_* = \mathbf{0}$ ensures rank$(\mathbf{Z}'_*) =$ rank$(\mathbf{Z}''_*) \le k$, while keeping the objective value in (4) unchanged. Hence, an optimal solution is given by

$$\mathbf{Z}'_* = \begin{bmatrix} \left(\mathbf{\Sigma}^{-1}\grave{\mathbf{U}}\mathbf{I}_k\right)\left(\mathbf{I}_k^\top\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top\right) \\ \mathbf{0}_{(d-m)\times n} \end{bmatrix} \quad \text{and,} \quad \mathbf{Z}_* = \mathbf{V}\mathbf{Z}'_* = \left(\mathbf{V}\mathbf{I}_m\mathbf{\Sigma}^{-1}\grave{\mathbf{U}}\mathbf{I}_k\right)\left(\mathbf{I}_k^\top\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top\right). \quad (7)$$

Case $\mathbf{m} > \mathbf{d}$: Recalling that the full SVD of $\mathbf{X} \in \mathbb{R}^{m\times d}$ is $\mathbf{X} = \mathbf{U}\widetilde{\mathbf{\Sigma}}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{m\times m}$, $\widetilde{\mathbf{\Sigma}} \in \mathbb{R}^{m\times d}$, and $\mathbf{V} \in \mathbb{R}^{d\times d}$, in this case, the last $m-d$ rows of $\widetilde{\mathbf{\Sigma}}$ will be zero, i.e., $\overline{\mathbf{I}}_{m-d}^\top\widetilde{\mathbf{\Sigma}} = \mathbf{0}$. Denote $\mathbf{Z}' := \mathbf{V}^\top\mathbf{Z}$. This time, as $\mathbf{\Sigma} := \mathbf{I}_d^\top\widetilde{\mathbf{\Sigma}}$,

$$\begin{aligned} \min_{\text{rank}(\mathbf{Z})\le k} \|\mathbf{X}\mathbf{Z} - \mathbf{Y}\|_{\text{F}}^2 &\equiv \min_{\text{rank}(\mathbf{Z}')\le k} \left\|\widetilde{\mathbf{\Sigma}}\mathbf{Z}' - \mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2 \\ &\equiv \min_{\text{rank}(\mathbf{Z}')\le k} \left\|\mathbf{\Sigma}\mathbf{Z}' - \mathbf{I}_d^\top\mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2 + \left\|\overline{\mathbf{I}}_{m-d}^\top\mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2 \\ &\equiv \min_{\text{rank}(\mathbf{Z}'')\le k} \left\|\mathbf{Z}'' - \mathbf{I}_d^\top\mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2 + \left\|\overline{\mathbf{I}}_{m-d}^\top\mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2 \\ &\overset{\text{(i)}}{=} \sum_{i=k+1}^{m} \sigma_i^2\left((\mathbf{U}\mathbf{I}_d)^\top\mathbf{Y}\right) + \left\|(\mathbf{U}\overline{\mathbf{I}}_{m-d})^\top\mathbf{Y}\right\|_{\text{F}}^2, \end{aligned} \quad (8)$$

where $\mathbf{Z}'' := \mathbf{\Sigma}\mathbf{Z}' = \mathbf{\Sigma}\mathbf{V}^\top\mathbf{Z}$. Note that the term $\left\|(\mathbf{U}\overline{\mathbf{I}}_{m-d})^\top\mathbf{Y}\right\|_{\text{F}}^2$ is the irreducible error, and (i) is, once again, a consequence of the fact that $\left\|\mathbf{Z}' - \mathbf{I}_d^\top\mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2$ is minimized by taking the best rank-k approximation of $\mathbf{I}_d^\top\mathbf{U}^\top\mathbf{Y}$. Moreover, since the SVD of $\mathbf{I}_d^\top\mathbf{U}^\top\mathbf{Y}$ is $\grave{\mathbf{U}}\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top$,

$$\begin{aligned} \mathbf{Z}''_* &:= \arg\min_{\text{rank}(\mathbf{Z}'')\le k} \left\|\mathbf{Z}'' - \mathbf{I}_d^\top\mathbf{U}^\top\mathbf{Y}\right\|_{\text{F}}^2 = \left(\grave{\mathbf{U}}\mathbf{I}_k\right)\left(\mathbf{I}_k^\top\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top\right), \\ \text{and} \quad \mathbf{Z}_* &= \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{Z}''_* = \left(\mathbf{V}\mathbf{\Sigma}^{-1}\grave{\mathbf{U}}\mathbf{I}_k\right)\left(\mathbf{I}_k^\top\grave{\mathbf{\Sigma}}\grave{\mathbf{V}}^\top\right). \end{aligned} \quad (9)$$

Additional Remarks

It is not necessary for $\mathbf{X}$ to be full rank. An equivalent result can be derived with $\mathbf{\Sigma}^{-1}$ replaced by $\mathbf{\Sigma}^{\dagger}$. Arriving at the globally optimal solution in lemma requires computing two SVDs, namely $\mathbf{X} \in \mathbb{R}^{m\times d}$, which entails a complexity of $\mathrm{O}(dm^2)$, and $\mathbf{U}^\top \mathbf{Y} \in \mathbb{R}^{m\times n}$, with a complexity of $\mathrm{O}(nm^2)$. Hence, the total computational complexity is $\mathrm{O}(m^2(n+d))$.

A simplification for Positive Definite Hessians

The expressions in (2) is a little involved, but it can be simplified if the Hessian, defined as $\mathbf{H} = \mathbf{X}^\top \mathbf{X} \in \mathbb{R}^{d\times d}$ is positive definite, i.e., all eignevalues are strictly greater than 0. Let $\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ be the eigenvalue decomposition of $\mathbf{H}$, and let $\mathbf{H}^{\frac{1}{2}} = \mathbf{U}\Lambda^{\frac{1}{2}}\mathbf{U}^\top$ be its symmetric square root. Furthermore assume that $\mathbf{H}$ is positive definite, i.e., all diagonal entries of $\mathbf{\Lambda}$ are strictly positive. This can be ensured by regularizing the Hessian a by adding an identity matrix to it, scaled by a small amount.

Then, consider the following equivalent optimization problems:

$$\begin{aligned}
\min_{\mathrm{rank}(\mathbf{Z})\le k} \left\| (\mathbf{A}-\mathbf{Z})\mathbf{X}^\top \right\|_{\mathrm{F}}^2 &\equiv \min_{\mathrm{rank}(\mathbf{Z})\le k} \mathrm{Tr}\left((\mathbf{A}-\mathbf{Z})\mathbf{H}(\mathbf{A}-\mathbf{Z})^\top \right) \\
&\equiv \min_{\mathrm{rank}(\mathbf{Z})\le k} \left\| (\mathbf{A}-\mathbf{Z})\mathbf{H}^{\frac{1}{2}} \right\|_{\mathrm{F}}^2 \\
&\equiv \min_{\mathrm{rank}(\mathbf{Z})\le k} \left\| (\mathbf{A}-\mathbf{Z})\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} \right\|_{\mathrm{F}}^2 \\
&\equiv \min_{\mathrm{rank}(\mathbf{Z})\le k} \left\| \mathbf{Y}-\mathbf{Z}\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}} \right\|_{\mathrm{F}}^2 \qquad (\text{where } \ \mathbf{Y} \triangleq \mathbf{A}\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}) \\
&\equiv \min_{\mathrm{rank}(\mathbf{Z}')\le k} \left\| \mathbf{Y}-\mathbf{Z}' \right\|_{\mathrm{F}}^2 \qquad (10)
\end{aligned}$$

Here, $\mathbf{Z}' \triangleq \mathbf{Z}\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}$, and the constraint $\mathrm{rank}(\mathbf{Z}') \le k$ follows from the fact that multiplying $\mathbf{Z}$ by an invertible matrix, $\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}$, keeps its rank unchanged. The final optimization problem can be solved optimally by considering the SVD of $\mathbf{Y}$ as $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, and the optimal solution is given by,

$$\mathbf{Z}'_* = (\mathbf{U}\mathbf{I}_k)\left(\mathbf{I}_k^\top \mathbf{\Sigma}\mathbf{V}^\top\right), \qquad (11)$$

where $\mathbf{I}_k$ denotes the first k columns of the identity matrix. So the optimal solution is given by,

$$\mathbf{Z}_* = (\mathbf{U}\mathbf{I}_k)\left(\mathbf{I}_k^\top \mathbf{\Sigma}\mathbf{V}^\top \mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{U}^\top\right). \qquad (12)$$

This gives a closed form expression for the optimal solution of the rank-constrained regression problem for the case of positive definite Hessians.

References

[1] Shuo Xiang, Yunzhang Zhu, Xiaotong Shen, and Jieping Ye. Optimal exact least squares rank minimization. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '12, page

480–488, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450314626. doi: 10.1145/2339530.2339609. URL `https://doi.org/10.1145/2339530.2339609`.

[2] Rajarshi Saha, Naomi Sagan, Varun Srivastava, Andrea J. Goldsmith, and Mert Pilanci. Compressing large language models using low rank and low precision decomposition, 2024. URL `https://arxiv.org/abs/2405.18886`.