Abbasnejad, Iman
(2017)
Rank minimization and sparse modeling.
(Unpublished)

Tech Report: Rank Minimization and Sparse Modeling

Iman Abbasnejad

In this document we provide more details on the presented bounds in Eq.8, Eq.9, Eq.10 and Eq.11 and also the relation between the ℓ_2-norm and nuclear norm in our paper [Abbasnejad et al.2017]. In this document we use the same notations and definitions as the paper.

1 Details of Equations 8 - 11

Lemma 4.1 suggests that the inner product between the linear combinations of two arbitrary $k-$sparse vectors $\mathbf{x}, \acute{\mathbf{x}}$ is approximately preserved by linear projection:

$$(1+\epsilon)\mathbf{x}^T\acute{\mathbf{x}} - 2R^2\epsilon \leq (A\mathbf{x})^T(A\acute{\mathbf{x}}) \leq (1-\epsilon)\mathbf{x}^T\acute{\mathbf{x}} + 2R^2\epsilon \tag{1}$$

by substituting $\|\mathbf{x}\|_2 \leq R, \|\acute{\mathbf{x}}\|_2 \leq R$ in the above equation we have (for more details see Lemma 4.2 and Lemma 4.3 in [Calderbank, Jafarpour, and Schapire2009]):

$$(1+\epsilon)\mathbf{x}^T\acute{\mathbf{x}} - (\|\mathbf{x}\|_2^2 + \|\acute{\mathbf{x}}\|_2^2)\epsilon \leq (A\mathbf{x})^T(A\acute{\mathbf{x}}) \leq (1-\epsilon)\mathbf{x}^T\acute{\mathbf{x}} + (\|\mathbf{x}\|_2^2 + \|\acute{\mathbf{x}}\|_2^2)\epsilon \tag{2}$$

by generalizing Eq. 2 to any arbitrary $k-$sparse vector $\mathbf{x}_i, \acute{\mathbf{x}_j}$ and substituting $\alpha_i y_i \mathbf{x}_i$ and $\acute{\alpha}_j \acute{y}_j \acute{\mathbf{x}}_j$ in Eq. 2 we have:

$$\begin{aligned}(1+\epsilon)\alpha_i\acute{\alpha}_j y_i\acute{y_j}\mathbf{x}_i^T.\acute{\mathbf{x}_j} - (\|\alpha_i y_i\mathbf{x}_i\|_2^2 + \|\acute{\alpha}_j\acute{y_j}\acute{\mathbf{x}}_j\|_2^2)\epsilon \\ \leq (A(\alpha_i y_i\mathbf{x}_i))^T(A(\acute{\alpha}_j\acute{y}_j\acute{\mathbf{x}_j})) \leq \\ (1-\epsilon)\alpha_i\acute{\alpha}_j y_i\acute{y_j}\mathbf{x}_i^T.\acute{\mathbf{x}_j} + (\|\alpha_i y_i\mathbf{x}_i\|_2^2 + \|\acute{\alpha}_j\acute{y_j}\acute{\mathbf{x}}_j\|_2^2)\epsilon\end{aligned} \tag{3}$$

Since linear SVM classifier is the linear combination of training examples:

$$\boldsymbol{\omega} = \sum_{i=1}^{M}\alpha_i y_i\mathbf{x}_i, \quad \acute{\boldsymbol{\omega}} = \sum_{j=1}^{N}\acute{\alpha_j}\acute{y_j}\acute{\mathbf{x}}_j \tag{4}$$

by getting summation over $\alpha_i y_i \mathbf{x}_i$ and $\acute{\alpha_j}\acute{y}_j\acute{\mathbf{x}}_j$ and substituting $\boldsymbol{\omega} = \sum_{i=1}^{M}\alpha_i y_i\mathbf{x}_i, \acute{\boldsymbol{\omega}} = \sum_{j=1}^{N}\acute{\alpha}_j\acute{y}_j\acute{\mathbf{x}}_j$ and $\mathbf{x}_i = \acute{\mathbf{x}}_j, N = M$ in Eq. 3 we have:

$$(1+\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} - 2\epsilon\sum_{i=1}^{M}\|\alpha_i y_i\mathbf{x}_i\|_2^2 \leq (A(\boldsymbol{\omega})^T(A(\acute{\boldsymbol{\omega}}) \leq (1-\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} + 2\epsilon\sum_{i=1}^{M}\|\alpha_i y_i\mathbf{x}_i\|_2^2 \tag{5}$$

from norm inequalities and definitions we know:

$$\sum_{i=1}^{M} \|\alpha_i y_i \mathbf{x}_i\|_2^2 \leq \sum_{i=1}^{M} \|\alpha_i y_i \mathbf{x}_i\|_2 \sum_{i=1}^{M} \|\alpha_i y_i \mathbf{x}_i\|_2 = \sum_{i=1}^{M} |y_i| \|\alpha_i \mathbf{x}_i\|_2 \sum_{i=1}^{M} |y_i| \|\alpha_i \mathbf{x}_i\|_2 = K^2 \sum_{i=1}^{M} \|\alpha_i \mathbf{x}_i\|_2 \sum_{i=1}^{M} \|\alpha_i \mathbf{x}_i\|_2 \quad (6)$$

where $\sum_{i=1}^{M} |y_i| \leq K$. Therefore we can rewrite Eq.5 as follows:

$$(1+\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} - 2K^2\epsilon \sum_{i=1}^{M} \|\alpha_i \mathbf{x}_i\|_2 \sum_{i=1}^{M} \|\alpha_i \mathbf{x}_i\|_2 \leq (A(\boldsymbol{\omega})^T(A(\acute{\boldsymbol{\omega}}) \leq (1-\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} + 2K^2\epsilon \sum_{i=1}^{M} \|\alpha_i \mathbf{x}_i\|_2 \sum_{i=1}^{M} \|\alpha_i \mathbf{x}_i\|_2 \quad (7)$$

following the same procedure and substituting $\boldsymbol{\omega} = \sum_{i=1}^{M} \beta_i y_i \mathbf{v}_i$ in Eq. 3 we have:

$$(1+\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} - 2K^2\epsilon \sum_{i=1}^{M} \|\beta_i \mathbf{v}_i\|_2 \sum_{i=1}^{M} \|\beta_i \mathbf{v}_i\|_2 \leq (A(\boldsymbol{\omega})^T(A(\acute{\boldsymbol{\omega}}) \leq (1-\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} + 2K^2\epsilon \sum_{i=1}^{M} \|\beta i \mathbf{v}_i\|_2 \sum_{i=1}^{M} \|\beta_i \mathbf{v}_i\|_2 \quad (8)$$

In order to compare the bounds in Eq. 8 and Eq. 10 we only need to compare:

$$\sum_{i=1}^{M} \|\alpha_i \mathbf{x}_i\|_2, \quad \sum_{i=1}^{M} \|\beta_i \mathbf{v}_i\|_2$$

To do so, from the definition we have:

$$\sum_{i=1}^{M} \|\alpha_i \mathbf{x}_i\|_2 \sum_{i=1}^{M} \|\alpha_i \mathbf{x}_i\|_2 = \sum_{i=1}^{M} |\alpha_i| \|\mathbf{x}_i\|_2 \sum_{i=1}^{M} |\alpha_i| \|\mathbf{x}_i\|_2$$
$$\sum_{i=1}^{M} \|\beta_i \mathbf{v}_i\|_2 \sum_{i=1}^{M} \|\beta_i \mathbf{x}_i\|_2 = \sum_{i=1}^{M} |\beta_i| \|\mathbf{v}_i\|_2 \sum_{i=1}^{M} |\beta_i| \|\mathbf{v}_i\|_2$$

and since $\sum_{i=1}^{M} |\alpha_i| = 1, \sum_{i=1}^{M} |\beta_i| = 1, \sum_{i=1}^{M} \|\mathbf{x}_i\| = \|\mathbf{X}\|_F, \sum_{i=1}^{M} \mathbf{v}_i = \|\mathbf{V}\|_F$, therefore we have:

$$(1+\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} - 2K^2\epsilon\|\mathbf{X}\|_F^2 \leq (A(\boldsymbol{\omega})^T(A(\acute{\boldsymbol{\omega}}) \leq (1-\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} + 2K^2\epsilon\|\mathbf{X}\|_F^2 \quad (9)$$

and:

$$(1+\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} - 2K^2\epsilon\|\mathbf{V}\|_F^2 \leq (A(\boldsymbol{\omega})^T(A(\acute{\boldsymbol{\omega}}) \leq (1-\epsilon)\boldsymbol{\omega}^T\boldsymbol{\omega} + 2K^2\epsilon\|\mathbf{V}\|_F^2 \quad (10)$$

and in order to compare the bounds we need to compare the Frobenius norm of the set of input examples, $\mathbf{X}, \mathbf{V}$

2 ℓ_2-norm vs. Nuclear norm

In this section we study the effect of nuclear norm minimization on the ℓ_2-norm.

In this section our goal is to show if $\|\mathbf{V}\|_* \leq \|\mathbf{X}\|_*$, then $\|\mathbf{X}\|_2 \leq \|\mathbf{V}\|_2$ where $\mathbf{V} = \mathbf{X} \circ \tau$ and τ is a set of transformation that only applies on the indexes and makes the rank of $\mathbf{V}$ as small as possible. In other words we want to show if:

$$\acute{r} \leq r \Longrightarrow \|\mathbf{V}\|_* \leq \|\mathbf{X}\|_* \Longrightarrow \sum_{i=1}^{\acute{r}} \acute{\sigma}_i \leq \sum_{i=1}^{r} \sigma_i \tag{11}$$

then:

$$\|\mathbf{X}\|_2 \leq \|\mathbf{V}\|_2 \Longrightarrow \sigma_1 \leq \acute{\sigma}_1 \tag{12}$$

where $rank(\mathbf{X}) = r$, $rank(\mathbf{V}) = \acute{r}$ and σ_i, σ_i are the i-th singular values of $\mathbf{V}$ and $\mathbf{X}$.

In order to see the relation between the ℓ_2-norm and the nuclear norm we can rewrite Eq. 11 as follows:

$$\|\mathbf{X}\|_* = \sigma_1 + \sigma_2 + \ldots + \sigma_r = C, \quad \|\mathbf{V}\|_* = \acute{\sigma}_1 + \acute{\sigma}_2 + \ldots + \acute{\sigma}_r = C - \epsilon \tag{13}$$

where ϵ is the difference between $\|\mathbf{X}\|_*$ and $\|\mathbf{V}\|_*$. By substituting C from the left hand side in Eq. 13 to the right hand side in Eq. 13 we have:

$$\acute{\sigma}_1 + \acute{\sigma}_2 + \ldots + \acute{\sigma}_{\acute{r}} = \sigma_1 + \sigma_2 + \ldots + \sigma_r - \epsilon \tag{14}$$

For comparison, we consider the worst case scenario in which $\sigma_1 = \sigma_2 = \ldots = \sigma_r$ and $\acute{\sigma}_1 = \acute{\sigma}_2 = \ldots = \acute{\sigma}_{\acute{r}}$, therefore:

$$\acute{r}\acute{\sigma}_1 = r\sigma_1 - \epsilon \tag{15}$$

substituting $\acute{\sigma}_1 = k\sigma_1$ in the above equation yields:

$$\acute{r}k\sigma_1 = r\sigma_1 - \epsilon \Longrightarrow k = \frac{r\sigma_1 - \epsilon}{\acute{r}\sigma_1} \tag{16}$$

in order to see when $k \geq 1$ and the $\|\mathbf{X}\|_2 \leq \|\mathbf{V}\|_2$ we have:

$$r\sigma_1 - \epsilon \geq \acute{r}\sigma_1 \Longrightarrow \epsilon \leq (r - \acute{r})\sigma_1 \tag{17}$$

where $(r - \acute{r}) \geq 0$. Fig. 1 visualizes the ℓ_2-norm vs. the nuclear norm. The details of datasets and implementations can be found in [Abbasnejad et al.2015, Abbasnejad et al.2016, Abbasnejad and Teney2015]

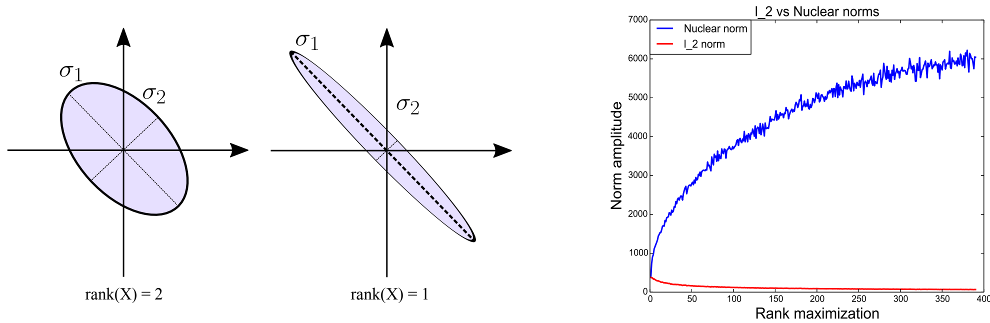

Figure 1: $\|\mathbf{X}\|_2$ vs $\|\mathbf{X}\|_*$ for different $r = rank(\mathbf{X})$

References

[Abbasnejad and Teney2015] Abbasnejad, I., and Teney, D. 2015. A hierarchical bayesian network for face recognition using 2d and 3d facial data. In *Machine Learning for Signal Processing (MLSP), 2015 IEEE 25th International Workshop on*, 1–6. IEEE.

[Abbasnejad et al.2015] Abbasnejad, I.; Sridharan, S.; Denman, S.; Fookes, C.; and Lucey, S. 2015. Learning temporal alignment uncertainty for efficient event detection. In *Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on*, 1–8. IEEE.

[Abbasnejad et al.2016] Abbasnejad, I.; Sridharan, S.; Denman, S.; Fookes, C.; and Lucey, S. 2016. Complex event detection using joint max margin and semantic features. In *Digital Image Computing: Techniques and Applications (DICTA), 2016 International Conference on*, 1–8. IEEE.

[Abbasnejad et al.2017] Abbasnejad, I.; Sridharan, S.; Denman, S.; ; Fookes, C.; and Lucey, S. 2017. From affine rank minimization solution to sparse modeling. In *Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on*, 1–7. IEEE.

[Calderbank, Jafarpour, and Schapire2009] Calderbank, R.; Jafarpour, S.; and Schapire, R. 2009. Compressed learning: Universal sparse dimensionality reduction and learning in the measurement domain. *preprint*.