Large scale semidefinite programming via saddle point mirror-prox algorithm

Z. Lu, A. Nemirovski, and R.D.C. Monteiro

This paper adapts the saddle point mirror-prox algorithm of Nemirovski to large scale SDPs with well-structured sparsity. The main results are in Section 3 on the economical representation of positive semidefiniteness of well-structured sparse symmetric matrices. The representation of a positive semidefinite matrix by the sum of smaller blocks of positive semidefinite matrices is of practical importance. More importantly, the authors showed that such a representation can be computed efficiently when the matrix has well-structured sparsity. In Section 4, the economical representation of the positive semidifiniteness of a matrix is applied to large scale SDPs with well-structured sparsity; in particular to SDP relaxations of the Lovasz capacity and MAX-CUT problems. Then in Section 5, the authors presented very impressive numerical results for the latter 2 classes of SDPs for problems with staircase sparsity structured. Some these SDPs have close to a million constraints and have matrix dimensions close to eighty thousands. They are far beyond the range that can be solved by interior-point methods or even first-order methods like SDPLR and spectral bundle method.

The content of the paper is excellent. The only criticism I may have is on the presentation of the results and the writing of the paper. The current version is very hard to read. I would urge the authors to think of ways to present the results in a more accessible manner, as well as improving the English. As it is now, some of the sentences are excessively long; many symbols and notations are embedded in the text; and there are numerous grammatical errors and typos.

I like the results established in Sections 3 and 4, and would like to see the paper get published. But I also hope to see a version that is better written than the current version.

Major comments and suggestions.

1. The paper concentrated on matrices with staircase sparsity structures. I think banded matrices are natural examples too. It would be nice to mention how you can apply the representation developed in Section 3 to a banded matrix. If one has a banded matrix, can the superiority of the mirror-prox algorithm over an interior-point algorithm in terms of arithmetic cost still be preserved?

2. The sparsity structured defined in Section 2 seems to bear some similarities to the skyline structure that is commonly used in numerical linear algebra to store sparse matrices. It would be worthwhile to mention the connections between these two concepts.

3. Bring the staircase example in p.9 to Section 2. It would help the reader to understand the notations in Section 2. Unless one sees an example, it is hard to understand the meaning of sets such as J_k, J'_k, etc. In fact, I would urge the authors to present Section 2 in a more readable manner.

Below are some typos that I found when reading the paper or clarifications needed in the paper. They are by no mean comprehensive.

[Abstract, last line] "are finally presented"

[p.1, (1)] $\cup$ should be $\cap$.

[p.1, ln -2] capable to produce high ...

Better to use "capable of producing high..."?

[p.2, middle] ... provided that the largest size μ of diagonal blocks in matrices from **S** ...

How about "... when the largest diagonal blocks in the matrices from **S** ..."?

[p.2, middle] ... the necessity to operate with eigenvalue decompositions .... rather than to be able to compute few largest eigenvalues of matrices from $\mathcal{N}$.

How about "the necessity to compute the eigenvalue decompositions .... rather than computing a few largest eigenvalues of the matrices in $\mathcal{N}$"?

The last two items are examples of sentences that are not easy to read because of the English. I hope the authors could work on the English to improve its readability.

[p.3, middle] from $\mathcal{N}$ is or is not positive semidefinite

"from $\mathcal{N}$ is positive semidefinite"

[p.3, bottom] "<<"

use the command `\ll`

[p.3, bottom] We are about to introduce terminology... allowing to operate with...

How about "Next, we introduce some terminology... for dealing with ..."

[p.3, ln -2] "$l \times l$ matrix with rows and columns indexed by elements from J

How about "the $l \times l$ matrix obtained from B by extracting the rows and columns with indices in J"

[p.4, top] "We refer to m as to"

Should be "We refer to m as the"

[p.4, item 4] "$\mathcal{J}_k$"

Should be "J_k"?

[p.4, item 6] Since $\mathbf{B}$ is used throughout the paper, it should be displayed more prominently in an equation, such as:

$$\mathbf{B} = \{B = (B_1, \dots, B_m) : B_k = [B^k_{ij} = B^k_{ji}]_{i,j \in J_k},\ k = 1, \dots, m\}$$

[p.4, item 7] "define linear mapping"

Should be "define the linear mapping"

[p.4, ln -1] How about "Finally, we let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of a symmetric matrix A, respectively"?

[p.5, ln 1] Should be "Consider again **the** semdifinite program (1). Assuming **that** the ..."

[p.6, ln 2] "if and only if it is **the** sum of".

[p.6, ln 3] Wrong font size, and missing "."

[p.6, middle] " $B^\epsilon \succeq 0 \succeq 0$"

[p.6, middle] "whence ... of the required structure"

Better to use "whence ... **has** the required structure"

[p.7, middle] "Vice versa"

Better to use "Conversely".

[p.7, middle] δ^i_j

Please mention that δ^i_j denotes the Kronecker delta.

[p.7, bottom] "due to $\mathcal{C}^\epsilon \succeq 0$"

Should be "$A^\epsilon \succeq 0$"? Please also define A^ϵ.

[p.8, ln 1] $\left[\Delta^k_{ij}\right]_{i,j} \in J'_{k+1}$

Should be "$\left[\Delta^k_{ij}\right]_{i,j \in J'_{k+1}}$"

[p.9, middle] "Collection d defines the subspace $\mathbf{S}^{[d]}$ of $d - staircase$ ..."

"Let $\mathbf{S}^{[d]}$ be the subspace of d-staircase ..."

[p.10, Section 3.2] "The cone $\mathbf{S}^{(v)}_+$ of ....."

Why not simply use:

Let $\mathbf{S}^{(v)}_+ = \mathbf{S}^{(v)} \cap \mathbf{S}^n_+$ and $\mathbf{C}^{(v)} = \{Z \in \mathbf{S}^{(v)} : \mathrm{Tr}(XZ) \geq 0 \ \forall\ X \in \mathbf{S}^{(v)}_+\}$?

[p.10, last line and p.11, ln 1] This line is not clear. I have a hard time trying to understand its meaning.

[p.12, bottom] How about

$$\begin{aligned} c^T x_\epsilon + \gamma c^T \bar{x} &\geq \mathrm{Opt} - \epsilon + T\lambda + \gamma c^T \bar{x} \\ &\geq \mathrm{Opt}(1+\gamma) - \epsilon \end{aligned}$$

where we have used (19) to get $T\lambda \geq \gamma(\mathrm{Opt} - c^T \bar{x})$.

[p.13, bottom] "at x. k-th"
Should be "at x. The k-th ...".

[p.14, top] "$(\mathbf{S}, \|\cdot\|)$"

Is $\|\cdot\|$ the matrix 2-norm?

[p.14, middle] "or, which is the same, is a rank 1 ..."

How about "or equivalently, is a rank 1 ..."?

[p.14, middle] $v = (v_1, v_1, \ldots, v_{n+1})^T$

Should be $v = (v_1, v_2, \ldots, v_{n+1})^T$?

[p.14, middle] "indexed $2, 3, \ldots, n+1$"

"indexed **by** $2, 3, \ldots, n+1$ "

[p.14, last line and p.15, ln 1–3] It is cleaner to define $\mathcal{M}$ in an equation, such as

$$\mathcal{M} = \left\{ \begin{bmatrix} \nu & \sqrt{\nu} e^T \\ \sqrt{\nu} e & Z \end{bmatrix} \in \mathbf{S}^{(v)} : Z_{ii} = \lambda \ \forall\, i = 1, \ldots, n,\ \lambda \in R,\ Z_{i-1,j-1} = 0 \text{ if } (i,j) \notin E \right\}$$

[p.16, middle] "Since $\|\Delta\|_{\mathcal{L}} = \|\mathcal{L}^{1/2} \Delta \mathcal{L}^{1/2}\| = \delta$, where δ is defined as in (32)"

[p.17, Prop 4.1] Then ϵ-solution to (29)

Should be "Then **an** ϵ-solution to (29)"

[p.18, bottom] Symbol like "$X^k(X)$" is quite confusing

Why not just use X^k?

[p.19, Remark 4.1] into the bound $N(\epsilon) \leq O(1) \frac{\mathrm{Opt}}{\epsilon}$

Why not use " $N(\epsilon) \leq O(1) \frac{\mathrm{Tr}(V)}{\epsilon}$"?

[p.19, bottom] ">>"

Please use the command \gg.

[p.20, middle] Is the computations done in C or MATLAB?