hw11 Optimization

Guillaume Jarry

November 2022

Remark: In this paper, it could happen that in describing the SRG of a class of operators I used the word circle instead of disc. I tried to correct this mistake as much as I could, but if you notice it, what I really meant was disc.

1 Problem 13.6

Let us notice that an intersection of SRG-full class operators is SRG full as well:

$$\mathcal{G}(1-\alpha(\mathcal{M}\cap\mathcal{L}_1) = 1-\alpha\mathcal{G}(\mathcal{M})\cap\mathcal{G}(\mathcal{L}_1)$$

where $\mathcal{G}(\mathcal{M})\cap\mathcal{G}(\mathcal{L}_1)$ is going to be the half circle of center 1, diameter α (and oriented toward the negative).

Thus it is not contained in $\mathcal{G}(\mathcal{L}_R)$ for $R<1$ and thus it is not possible to derive convergence from it.

Because of the SRG-fullness of the class, the containement of the SRG is equivalent to the containement of the class.

2 Problem 13.8

First of all, the resolvant of A will be SRG-full as, A is the intersection of two SRG-full class and that the operations involved in the computation of the SRG preserve SRG-fullness.

$$J_{\alpha A} = (I+A)^{-1}$$

Therefore, following the rule of computation of SRG for SRG full classes gives:

$$\mathcal{G}(J_{\alpha A}) = \mathcal{G}((I+A))^{-1}$$

$$\mathcal{G}(J_{\alpha A}) = (1+\alpha\mathcal{G}(A))^{-1}$$

$$\mathcal{G}(J_{\alpha A}) = (1 + \alpha\mathcal{G}(\mathcal{L}_\gamma^{-1}) \cap \mathcal{G}(\mathcal{M})))^{-1}$$

First, we notice that $\mathcal{G}(\mathcal{L}_\gamma)$ is the disc of center 0 and radius γ, therefore its inverse is $\bar{C}/C_{\frac{1}{\gamma}}$ where $C_{\frac{1}{\gamma}}$ is the disc of center zero and radius $\frac{1}{\gamma}$. Second, $\mathcal{G}(\mathcal{M}))$ is the set $\{a + ib | a \geq 0\}$. Then we have to take into acocunt the multiplication by factor α which act as an homothety of factorα center 1.

We finally have to compute the geometric inverse of the set:

$$\{a + ib | a \geq 1\}/C_{1,\frac{\alpha}{\gamma}} = \left\{ 1 + Re^{i\theta} | R > \frac{\alpha}{\gamma}, \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \right\}$$

Where $C_{1,\frac{1}{\gamma}}$ is the disc of center 1 and radius $\frac{1}{\gamma}$. This expression gives:

$$\mathcal{G}(J_{\alpha A}) = \left\{ Re^{i\theta} | \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}], R \leq \frac{\gamma}{\sqrt{\alpha^2 + \gamma^2 + 2\alpha^2 cos(\theta)}} \right\}$$

And we can simplify the right expression even more by removing the dependancy in θ (it becomes an nclusion then and ceases to be an equality) which gives the half disc of center zero and radius inferior to R, obviously included in the disk of same center and same radius:

$$\mathcal{G}(J_{\alpha A}) \subset \left\{ Re^{i\theta} | \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}], R \leq \frac{\gamma}{\sqrt{\alpha^2 + \gamma^2}} \right\}$$

Which proves the inclusion $\mathcal{G}(J_{\alpha A}) \subset \mathcal{G}(\mathcal{L}_R)$, but also, the tightness of the result. The value $R = \frac{\gamma}{\sqrt{\alpha^2+\gamma^2}}$ is the minimal value for which the inclusion happens.

3 Problem 3.11

First we toice that $\mathcal{L}_1$ and $\mathcal{A}$ being both SRG-full class, it results from theorem 26 that the composition is commutative ie:

$$\mathcal{G}(\mathcal{L}_1\mathcal{A}) = \mathcal{G}(\mathcal{A}\mathcal{L}_1) = \mathcal{G}(\mathcal{L}_1)\mathcal{G}(\mathcal{A})$$

Then, we have to compute the product $\mathcal{G}(\mathcal{L}_1)\mathcal{G}(\mathcal{A})$ where $\mathcal{G}(\mathcal{L}_1)$ is the disc of center 0 and radius 1. Then, because there exist an element z of $\mathcal{G}(\mathcal{A})$ whose module is R, the multiplication of that element with $\mathcal{G}(\mathcal{L}_1)$ gives $\mathcal{G}(\mathcal{L}_R)$.

Then, for all the other elements of $\mathcal{G}(\mathcal{A}) \subset \mathcal{G}(\mathcal{L}_R)$, their module is going to be inferior to R, thus their multiplication with $\mathcal{G}(\mathcal{L}_1)$ will also be included in $\mathcal{G}(\mathcal{L}_R)$.

$$\mathcal{G}(\mathcal{L}_1)\mathcal{G}(\mathcal{A}) = \cup_{a \in \mathcal{A}} a\mathcal{G}(\mathcal{L}_1) = \mathcal{G}(\mathcal{L}_R) \cup_{a \in \mathcal{A}/\{z\}} a\mathcal{G}(\mathcal{L}_1)$$

ie a union of a set with set that are included inside it. Thus:

$$\mathcal{G}(\mathcal{L}_1)\mathcal{G}(\mathcal{A}) = \mathcal{G}(\mathcal{L}_R)$$

Then, using the rules of computation of operators with the identity operator and the identity exposed above:

$$\mathcal{G}(\frac{1}{2}I + \frac{1}{2}\mathcal{A}\mathcal{L}_1) = \frac{1}{2} + \frac{1}{2}\mathcal{G}(\mathcal{A}\mathcal{L}_1) = \frac{1}{2} + \frac{1}{2}\mathcal{G}(\mathcal{L}_1\mathcal{A})$$

Using the commutativity of the product and rearranging the expression the same way we developped it:

$$\mathcal{G}(\frac{1}{2}I + \frac{1}{2}\mathcal{A}\mathcal{L}_1) = \mathcal{G}(\frac{1}{2}I + \frac{1}{2}\mathcal{L}_1\mathcal{A})$$

And finally too using the inclusion of averaged operators and Lipschitz-Operators;

$$\mathcal{G}(\frac{1}{2}I + \frac{1}{2}\mathcal{A}\mathcal{L}_1) = \frac{1}{2} + \frac{1}{2}\mathcal{G}(\mathcal{L}_R) = \mathcal{G}(\mathcal{N}_{\frac{1}{2}}) \subset \mathcal{G}(\mathcal{L}_{\frac{1}{2}+\frac{1}{2}R}$$

Which is the expected results and closes the problem.

4 Problem 3.12

Using the hint and the already completed proof of therorem 19, we get:

$$\mathcal{G}(\mathcal{M}_\mu) = \mu + \mathcal{G}(\mathcal{M}) = \{a + ib | a > \mu\}$$

Which is exactly the expected result for the SRG of μ-strongly monotone function.

Then, using the SRG-fullness of C_β and $\mathcal{M}_\mu$, we get the equality:

$$\mathcal{G}(\mathcal{C}_\beta) = \mathcal{G}(\mathcal{M}_\mu)^{-1} = \left\{\beta + Re^{i\theta} | R \geq 0, \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]\right\}$$

Re-using the computation of problem 3.8 and adapting them, we get the inverse:

$$\mathcal{G}(\mathcal{C}_\beta) = \left\{Re^{i\theta} | \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}], R \leq \frac{1}{\sqrt{\beta^2}} = \frac{1}{\beta}\right\}$$

And finally, for the θ-averaged operators, the rules of computation of SRG gives us:

$$\mathcal{G}(\mathcal{N}_\theta) = (1-\theta)\mathcal{G}(\mathcal{L}_\theta)$$

ie the circle of center $1-\theta$ and of diameter θ, which is exactly the expected result !

5 Problem 3.14

For $\mathcal{M}$, we notice that the image of R^2 by the function $f_{\mathcal{M}}(x,y) \to (|x|, y)$ (which is nonnegative homogeneous), is the SRG of $\mathcal{M}$.

6 Problem 3.20

In all this problem, we use the result that the resolvant of a SRG-full class is SRG full because the operation used in the computation of the resolvant do not impact SRG-fullness.

a) Developping using the resolvant identity we get:

$$\mathcal{G}(R_{\mathcal{B}}) = 2\mathcal{G}(1+\mathcal{C}_{\beta})^{-1} - 1$$

We will use the characterisation of theorem 19 to more efficiently compute the inverse of $\mathcal{G}(I+\mathcal{C}_{\beta}) = \{1+z \in C | Re(z) \geq \beta |z|^2\}$. Knowing that for any complex number z, $z^{-1} = \frac{\bar{z}}{|z|^2}$. Therefore the inverse becomes:

$$\mathcal{G}(1+\mathcal{C}_{\beta})^{-1} =$$

And then we have to remind ourselves that: $\mathcal{G}(\mathcal{N}_{\frac{1}{1+\beta}})$ will be a disc of center $\frac{\beta}{1+\beta}$ and radius $\frac{1}{1+\beta}$. Therefore the equality stands and the equality of the SRG of two SRG full class is equivalent to the equality of the classes and thus:

$$R_{\mathcal{B}} = \mathcal{N}_{\frac{1}{1+\beta}}$$

b) Then, using the fact that $1+\mathcal{M}_{\mu} = \mathcal{M}_{1+\mu}$, and that $\mathcal{G}(\mathcal{M}_{1+\mu})^{-1} = \mathcal{G}(\mathcal{C}_{1+\mu})$ thanks to the problem 3.12, we get:

$$\mathcal{G}(-R_{\mathcal{A}}) = -2\mathcal{G}(1+\mathcal{M}_{\mu})^{-1} + 1 = -2\mathcal{G}(\mathcal{M}_{1+\mu})^{-1} + 1 = 1 - 2\mathcal{G}(\mathcal{C}_{1+\mu})$$

ie the disk of center $1 - \frac{1}{1+\mu} = \frac{\mu}{1+\mu}$ and of diameter $\frac{1}{1+\mu}$. And that is exactly the SRG of $\mathcal{N}_{\frac{1}{1+\mu}}$

As usual, the equality of the SRG of the two class guaranties the equality of the class for two class of SRG full operators, therefore:

$$R_{\mathcal{A}} = \mathcal{N}_{\frac{1}{1+\mu}}$$

c) Using the two equalities exposed aboved, (and the SRG-fullness of the two class), we get:

$$-R_{\mathcal{A}}R_{\mathcal{B}} = \mathcal{N}_{\frac{1}{1+\beta}}\mathcal{N}_{\frac{1}{1+\mu}}$$

which, given theorem 27, simplifies into a averaged operators of coefficient

$$\frac{\theta_1 + \theta_2 - 2\theta_1\theta_2}{1 - \theta_1\theta_2} = \frac{\mu + \beta}{\mu + \beta + \mu\beta}$$

which in this case $\theta_1 = \frac{1}{1+\mu}$ and $\theta_2 = \frac{1}{1+\beta}$. This results in:

$$-R_{\mathcal{A}} R_{\mathcal{B}} \subset \mathcal{N}_{\frac{\mu+\beta}{\mu+\beta+\mu\beta}}$$

d) Then, $\frac{1}{2}$-averaging a -1 averaged operator doesn't do anything. More precisely, using the result of problem 13.2

$$\frac{1}{2} + \frac{1}{2} R_{\mathcal{A}} R_{\mathcal{B}} = \frac{1}{2} I - \frac{1}{2} \mathcal{N}_{\frac{\mu+\beta}{\mu+\beta+\mu\beta}} = \frac{1}{2} \mathcal{C}_{\frac{1}{2\theta}}$$

Where $\theta = \frac{\mu+\beta}{\mu+\beta+\mu\beta}$. ie, when applying the SRG, because of the multiplication by $\frac{1}{2}$ we get a disk of center $1-\theta$ and of radius θ, which s exactly the SRG of the class of θ averaged operator. Therefore, (because of the SRG fullness of the classes) we get:

$$\frac{1}{2} + \frac{1}{2} R_{\mathcal{A}} \subset \mathcal{N}_{\frac{\mu+\beta}{\mu+\beta+\mu\beta}}$$

e)

7 Problem 4.1

Let us consder the following algorithm where we precompute $(I + \alpha A^T A)^{-1}$ using the matrix inversion lemma and noticing that $I + \alpha A^T A$ is actually a symetric matrix, therefore the inversion can be computed by singular value decomposition whose complexity is $\mathcal{O}(m^2 n)$ (which is an improvement over the classical $\mathcal{O}(n^3)$ inversion). Once we do that at the begining, then the cost of the iteration become the cost of the multiplication of a matrix with one vector, which is $\mathcal{O}(mn)$ (note that the vector $x + \alpha b$ takes $2n$ flops to compute).

So in the end, we have an algorthm that requires a precompute time of $\mathcal{O}(m^2 n)$ and a cost per iteration of $\mathcal{O}(mn)$, which is cheaper than to compute the inverse ($\mathcal{O}(n^3)$) at each iteration.

8 Problem 4.2

To solve this problem, we will apply a consensus technique. The problem is equivalent to solving:

$$\underset{x \in C}{minimize} \sum_{i=1}^{l} g_i(A_i x_i)$$

Where $C = \{(x_1, \ldots, x_n) | x_1 = \ldots x_n\} \subset R^{m \times l}$, and where the projection onto the consensus set is simple averaging $\bar{x} = \frac{1}{l} \sum_{i=1}^{l} x_i$. We can then apply the DRS:

$$\forall i \in \{1, \ldots, n\}, \qquad x_i^{k+1} = Prox_{\alpha g \circ A}(2\bar{z}^k - z_i^k)$$

$$z^{k+1} = z^k + x^{k+1} - \bar{z}^k$$

With $\circ$ as the composition symbol between g_i and the matrix multiplication A_i

The computation cost of the matrix $A_i x$ is $\mathcal{O}(mn)$, the cost of a $Prox_g$, $\mathcal{O}(C_g)$. Therefore, because we have l distinct proxima and we need to compute the matrix vector product beforehand, we correctly get a computaton cost of $\mathcal{O}(lmn + lC_g)$ flops. The last line havng a computation cost of $\mathcal{O}(n)$ it does not change the overall complexity exposed above. Therefore, parralellizing the computation of the matrices and the proxima (by assigning them to l different processors) could divide the complexity of the algorithm by l, and thus reach a complexity of $\mathcal{O}(mn + C_g)$, which is much better ! The consensus technique fits parrallelization quite well.