Scaled Relative Graph

Robert Hannah · Ernest K. Ryu · Wotao Yin

October 10, 2016

1 Introduction

This paper introduces the Scaled Relative Graph (SRG): an intuitive and powerful tool that radically simplifies the monotone operator theory and some convex analysis. It creates a two-dimensional graphical signature of an operator's action, which explain the convergence behaviors of many algorithms in convex optimization via a simple, unified, rigorous approach. It provides visual intuitions behind many existing theorems about operators and makes it easy to prove new theorems (as this paper will do).

Sitting at a high level of generality, monotone operator theory abstracts the components of a convex optimization algorithm as operators and relates the convergence behavior of the algorithm to the properties of these operators. However, this extremely useful theory has a high barrier to entry as the definitions and results are expressed in terms of norms, inner products, and inequalities, which can be unintuitive.

The SRG makes monotone operator theory much more intuitive. To illustrate the difficulty, consider the following result.

Theorem 1 *Let $\mathbb{H}$ be a Hilbert space. Let operator $C : \mathbb{H} \to \mathbb{H}$ be σ-strongly monotone. Then its Yoshida approximation ${}^{\gamma}C$ is $1/\gamma$-cocoercive, and $\frac{1}{2}\left(1 - \frac{\sigma}{1+\sigma\gamma}\right)$-averaged.*

To the best of our knowledge, this result is found in the existing literature. Even for readers already familiar with the main definitions, Theorem 1 is likely unintuitive and non-obvious. With the SRG, it become intuitive and obvious.

The operators in Theorem 1 naturally arise in convex optimization, and their assumed properties follow from the properties of convex functions and sets. Operator compositions like $B \circ A$ in part 1 arise in operator splitting, a technique that reduces a problem into simpler subproblems. The Yoshida approximation in part 2 and the reflected resolvent in part 3 appear in proximal algorithms, and the latter is a component of the so-called Douglas-Rachford and ADMM algorithms. Projections are used for enforcing constraints. All the results in Theorem 1 are important to understanding related algorithms and deriving their convergence rates. Although they appear to be very abstract, they will become straightforward with SRGs. Therefore, the SRG is also a tool for analyzing the algorithms for convex optimization.

Department of Mathematics, University of California, Los Angeles, CA 90095, USA

The SRG can be thought of as a two-dimensional visual signature of an operator (much like the spectrum of a linear operator). We take an operator on a general Hilbert space $\mathbb{H}$, and reduce it to a two-dimensional graph that contains the most important information about it. The shape of the graph encodes many properties of the operator including:

- Whether it is Lipschitz, monotone, averaged, cocoercive, etc.
- The real spectrum for a linear operator;
- Whether it is injective, multi-valued, etc.

The SRG contains enough information to allow you to rigorously prove theorems with simple geometry as it provides equivalent statements relating an operator to its SRG. Transformations of the operator correspond to transformations of the graph. Because of this, it is possible to chain implications together and prove general theorems about operators.

SRG is based off ideas found in [9, 10, 12, 11], where the authors maps the action of single-valued functions $A : \mathbb{R}^2 \to \mathbb{R}^2$ to $\mathbb{R}^2$ and use it as an illustration. In contrast, our SRG is developed for single- and multi-valued operators on a general Hilbert space. Moreover, we use the SRG not just for illustration, but rather as a rigorous tool for proofs based on chains of "if and only if" statements.

2 Operators

In this section, we discuss preliminaries and set up the notation. We refer readers to standard references for more information on convex analysis [13, 8, 7], nonexpansive and monotone operators [6, 18], and geometry [19, 16, 17].

Write $\mathbb{H}$ for a real Hilbert space equipped with the inner product $\langle \cdot, \cdot \rangle$. Given $\alpha \in \mathbb{R}$ and sets $U, V \subseteq \mathbb{H}$, write

$$\alpha U = \{\alpha u \,|\, u \in U\}, \quad U + V = \{u + v \,|\, u \in U,\, v \in V\}, \quad U - V = U + (-V).$$

Notice that if either U or V is $\emptyset$, then $U + V = \emptyset$ and $U - V = \emptyset$.

Multi-valued operators. Multi-valued operators, which map a point to a set, is a very natural notion in analyzing nonexpansive operators as there is a one-to-one correspondence between (single-valued) nonexpansive operators and (multi-valued) monotone operators. In particular, A is nonexpansive if and only if there exists a multi-valued monotone operator B such that $A = 2J_B - I$. We define monotonicity and the resolvent J_B soon.

We say A is a (multi-valued) operator on $\mathbb{H}$ and write $A \colon \mathbb{H} \rightrightarrows \mathbb{H}$ if A maps a point in $\mathbb{H}$ to a (possibly empty) subset of $\mathbb{H}$. So $A(x) \subseteq \mathbb{H}$ for all $x \in \mathbb{H}$. For simplicity, we write $Ax = A(x)$. Define $\mathrm{dom}(A) = \{x \,|\, Ax \neq \emptyset\}$. If $A : \mathbb{H} \rightrightarrows \mathbb{H}$ always maps a point to a singleton or the empty set, we say A is single-valued and identify it with the function $\tilde{A} : \mathrm{dom}(A) \to \mathbb{H}$ such that $\{\tilde{A}(x)\} = A(x)$ for all $x \in \mathrm{dom}(A)$. Given a multi-valued operator A, we say that A' is a selection of A if A' is single-valued, and $A'x \subset Ax$ for all x. We let $\mathcal{S}(A)$ be the set of selections of A.

The graph of an operator is defined as

$$\mathrm{graph}(A) = \{(x, u) \,|\, u \in Ax\}.$$

For convenience, we do not distinguish an operator from its graph, usually writing $(x, u) \in A$ to mean $u \in Ax$. Write $I : \mathbb{H} \to \mathbb{H}$ for the identity operator on $\mathbb{H}$. Define the inverse operator as

$$A^{-1} = \{(u, x) \,|\, (x, u) \in A\},$$

which always exists. This is not an inverse in the usual sense since $A^{-1}A \neq I$ is possible. The resolvent J_A of A is defined as $(I+A)^{-1}$. Given $A : \mathbb{H} \rightrightarrows \mathbb{H}$, write αA and $A\alpha$ for the operators respectively defined by $(\alpha A)(x) = \alpha(Ax)$ and $(A\alpha)(x) = A(\alpha x)$ for all $x \in \mathbb{H}$. To clarify, $\alpha(Ax)$ is a scalar-set product (defined above). We respectively call αA and $A\alpha$ post and pre-scalar multiplication of the operator A. Given $A : \mathbb{H} \rightrightarrows \mathbb{H}$ and $B : \mathbb{H} \rightrightarrows \mathbb{H}$, define $A+B : \mathbb{H} \rightrightarrows \mathbb{H}$ with $(A+B)(x) = Ax + Bx$. To clarify, $Ax+Bx$ is the sum of sets. Define $A-B$ similarly. Note that $\mathrm{dom}(A+B) = \mathrm{dom}(A) \cap \mathrm{dom}(B)$, since $\emptyset + W = \emptyset$ for any set W. Likewise, define the composition $A \circ B$ with

$$(A \circ B)(x) = \cup_{u \in Bx} Au.$$

We often omit $\circ$ and simply write AB to mean $A \circ B$. AB has domain $A^{-1}(\mathrm{dom}(B))$.

For $L \in (0,\infty)$, we say an operator is L-Lipschitz if

$$\|Ax - Ay\|^2 \leq L^2 \|x-y\|^2 \qquad \forall\, x,y \in \mathbb{H}.$$

We call an operator a contraction if it is L-Lipschitz for some $L < 1$, and nonexpansive if it is 1-Lipschitz. For $\beta \in (0,\infty)$, we say an operator is β-cocoercive if

$$\langle Ax - Ay, x-y \rangle \geq \mu \|Ax - Ay\|^2 \qquad \forall\, x,y \in \mathbb{H}.$$

An operator is called *firmly nonexpansive* if it is 1-cocoercive. We say an operator is monotone if

$$\langle Ax - Ay, x-y \rangle \geq 0 \qquad \forall\, x,y \in \mathbb{H}.$$

To clarify, $\langle Ax - Ay, x-y \rangle \geq$ means $\langle u-v, x-y \rangle \geq 0$ for all $(x,u),(y,v) \in A$. If $x \notin \mathrm{dom}(A)$, then the inequality is vacuous. For $\mu \in (0,\infty)$, we say an operator is μ-strongly monotone if

$$\langle Ax - Ay, x-y \rangle \geq \mu \|x-y\|^2 \qquad \forall\, x,y \in \mathbb{H}.$$

Monotone operators and strongly-monotone operators can be multi-valued. For $\theta \in (0,1)$, we say an operator T is θ-averaged if T that can be written as

$$T = (1-\theta)I + \theta S$$

for S that is nonexpansive.

Convex analysis. A function $f : \mathbb{H} \to \mathbb{R} \cup \{\infty\}$ is convex if

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta) f(y), \qquad \forall\, x,y \in \mathbb{H},\, \theta \in (0,1).$$

We say f is proper if $f(x) < \infty$ somewhere. Let let $\mathit{\Gamma}_0(\mathbb{H})$ denote the class of proper lower semicontinuous, convex functions defined on $\mathbb{H}$. We define the *domain* of f as the set of points on which f is finite: $\mathrm{dom}(f) = \{x | f(x) < \infty\}$. We say f is μ-strongly convex if $f(x) - (\mu/2)\|x\|^2$ is convex and L-smooth if ∇f is L-Lipschitz continuous. Write

$$\partial f(x) = \{g \in \mathbb{H} \,|\, f(y) \geq f(x) + \langle g, y-x \rangle,\, \forall y \in \mathbb{H}\}$$

for the subdifferential of a convex function f at x.

Inversive geometry. Consider the extended 2D plane $\overline{\mathbb{R}^2} = \mathbb{R}^2 \cup \{\infty\}$. We call the function $V : \overline{\mathbb{R}^2} \to \overline{\mathbb{R}^2}$, the *inversion* map.

$$V(v) = \frac{1}{\|v\|} v$$

where $V(0) = \infty$ and $V(\infty) = 0$. V a bijection, and is well-defined at $0, \infty$. In complex analysis, the inversion map is known as the Möbius transformation [1, p. 366]. More generally, inversive geometry is a classical tool for solving problems in Euclidean geometry [17, p. 75].

Fixed-point iterations. Given $T : \mathbb{H} \to \mathbb{H}$, consider the fixed-point iteration given by

$$x^{k+1} = Tx^k$$

for $k = 0, 1, \ldots$ where $x^0 \in \mathbb{H}$ is a starting point. We say $x^\star$ is a *fixed point* of T if $x^\star = Tx^\star$. If an operator T is nonexpansive, the fixed-point iteration may not converge. For instance, if $T = -I$, then x^k oscillates between x^0 and $-x^0$. If T is a contraction with Lipschitz constant $L < 1$ then $x^k \to x^\star$ strongly to the fixed point $x^\star$ with rate $\mathcal{O}(L^k)$. This classical argument is the basis of the Banach contraction principle [3].

If T is averaged, i.e., $T = (1-\theta)I + \theta R$ for some nonexpansive operator R and $\theta \in (0, 1)$, then $x^k \to x^\star$ weakly for a fixed point $x^\star$ provided that T has a fixed point. This result is the basis of the Krasnosel'skiĭ–Mann iteration [15,14]. The assumption of averagedness is stronger than nonexpansiveness and weaker than contractiveness. (The latter fact is easy to see using the SRG.)

3 The Scaled Relative Graph

The SRG of an operator A is a region of the extended 2D plane $\mathcal{G}(A) \subset \overline{\mathbb{R}^2}$. This region acts as a signature, and captures the essential information necessary to rigorously prove theorems. Simple transformations of an operator, such as scaling, transition, and inversion lead to simple transformations of the SRG. In this section we define the SRG.

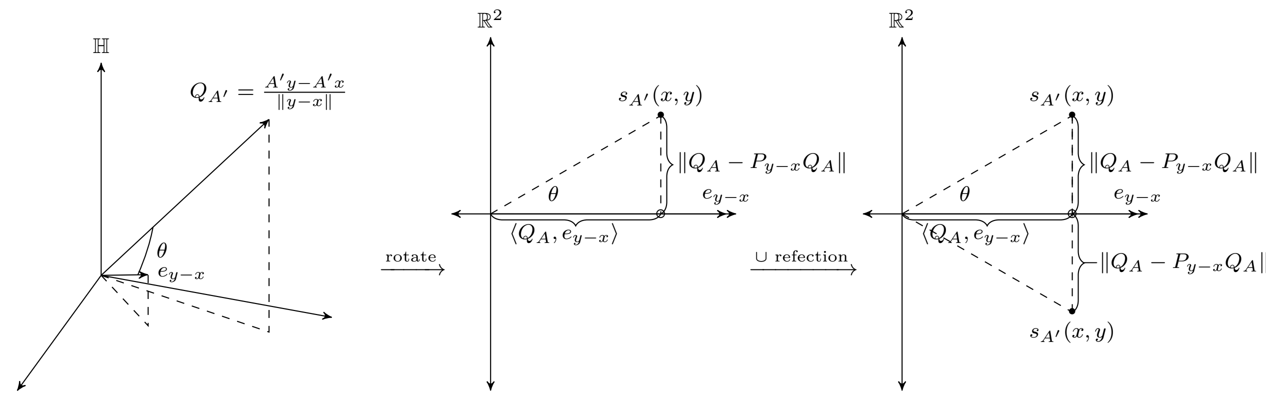
3.1 Construction and interpretation

We define the SRG in a number of steps. We define $\mathbb{P} \subset \mathbb{H}^2$ as the set of points of the form (x, x) for $x \in \mathbb{H}$.

1. **Quotient map:** The starting point is the quotient map $Q_A : \mathbb{H}^2 \backslash \mathbb{P} \rightrightarrows \mathbb{H}$. This we can think of as the "vector slope" of the operator A on the interval from x to y. As we travel from x to y, the value changes from Ax to Ay. We divide this change in A (which is $Ay - Ax$) by the distance over which we traveled (which is $\|y - x\|$) much like "rise over run."

$$Q_A(x, y) = \frac{Ay - Ax}{\|y - x\|}, \quad \text{for } x \neq y. \tag{1}$$

   where $Ay - Ax$ is interpreted as setwise subtraction defined in Section 2. Q_A is multi-valued in general because A might be.
2. **graphing map:** Using the quotient map $Q_A(x, y)$, we start drawing a graph in the complex plane. Let A' be single-valued.
   (a) On the real axis we plot the component of $Q_{A'}(x, y)$ in the direction of $y - x$, denoted as $e_{y-x} = \frac{y-x}{\|y-x\|}$. This is the direction along which we have traveled. This we can think of as the "parallel component" of the "vector slope".
   (b) Next, on the imaginary axis, we plot $\pm$ the component of $Q_{A'}$ that is perpendicular to $y - x$. We can think of this as the "normal component" of the "vector slope".



The complex number that we obtain we denote by $s_{A'}(x,y)$ is the graphing map. For a multi-valued operator A, $S_A(x,y)$ is just the union of all point $S_{A'}(x,y)$ for selections $A' \in \mathcal{S}(A)$. We now define this precisely with a number of equivalent definitions. Depending on context, some definitions are easier to apply or interpret than others.

graphing map: Given an operator $A : \mathbb{H} \to \mathbb{H}$, the *graphing map* $s_A : \mathbb{H}^2 \backslash \mathbb{P} \to \mathbb{R}^2$ is defined as follows:

$$s_A(x,y) = \bigcup_{A' \in \mathcal{S}(A)} \langle Q_{A'}(x,y), e_{y-x}\rangle \pm i\|Q_{A'}(x,y) - \mathrm{Proj}_{y-x}(Q_{A'}(x,y))\| \tag{2a}$$

$$= \bigcup_{A' \in \mathcal{S}(A)} \langle Q_{A'}(x,y), e_{y-x}\rangle \pm i\sqrt{\|Q_{A'}(x,y)\|^2 - \langle Q_{A'}(x,y), e_{y-x}\rangle^2} \tag{2b}$$

where $e_{y-x} = \frac{y-x}{\|y-x\|}$, and $\mathcal{P}_{y-x}$ is the projection onto $\mathrm{span}\{y-x\}$. Alternately, if θ is the angle between $A'y - A'x$ and $y - x$, then

$$s_A(x,y) = \bigcup_{A' \in \mathcal{S}(A)} \frac{\|A'y - A'x\|}{\|y-x\|}(\cos\theta, \pm\sin\theta) \tag{3a}$$

$$= \bigcup_{A' \in \mathcal{S}(A)} \|Q_{A'}(x,y)\|(\cos\theta, \pm\sin\theta). \tag{3b}$$

3. **Scaled Relative Graph:** Finally, the scaled relative graph (SRG) of an operator is the image of the operator under the graphing map. We also include the "point at infinity" $\{\infty\}$ if the operator A is multi-valued.

$$\mathcal{G}(A) = \bigcup_{x \neq y} s_A(x,y) \;\; (\cup\{\infty\} \text{ if } A \text{ is multi-valued}) \tag{4}$$

If an operator A is multi-valued at x, say, we can think of its slope as infinite there, because it made a nonzero jump in zero distance. The point at infinity has an important role in rigorously proving operator properties.

Here we discuss the interpretation of *individual points* on the SRG (leaving the interpretation of the *entire graph* to a later section):

1. **The magnitude of s_A:** is the average rate of change of A when moving from x to y:
$$\|s_{A'}(x,y)\| = \|Q_{A'}(x,y)\| = \|A'y - A'x\|/\|y-x\|$$
2. **The horizontal component of** s_A is how much of that change is in the direction of $y-x$, which is the direction we move;
3. **The vertical component of** s_A is how much of that change is perpendicular to the direction of $y-x$;
4. **The angle of incidence of** s_A tells the angle between the change in A (which is $Ay - Ax$) and the change in position (which is $y-x$). The closer this angle is to $\pi/2$, the more that A is twisting us away from our direction of motion. The closer this angle is to 0, the more that A is tending towards our direction of motion.
5. **The point** ∞ indicates that the operator is multi-valued at some point.
6. **The point** 0 indicates that the operator is not injective, since $0 \in \mathcal{G}(A)$ if and only if $Ax = Ay$ for $x \neq y$.
7. **Points on the horizontal axis** correspond exactly to the set of real eigenvalues of a linear operator.

3.2 Example Calculations of the Scaled Relative Graph

To clarify the definitions, we will calculate the SRGs of a few operators.

Example 1 (basic operators)

1. The SRG of I, the identity operator, consists of the single point $(1,0)$.
2. Let $c, d \in \mathbb{R}$. The SRG of $cI + d$, is just $(c, 0)$ regardless of d.
3. The SRG of R_θ, a rotation of θ radians in $\mathbb{R}^2$ is: $(\cos\theta, \pm\sin\theta)$.

Proof 1. Choose any $x \neq y$,
$$Q_I(x,y) = \frac{y-x}{\|y-x\|} = e_{y-x}$$
$$s_I(x,y) = \langle e_{y-x}, e_{y-x}\rangle \pm i\sqrt{\|e_{y-x}\|^2 - \langle e_{y-x}, e_{y-x}\rangle^2} = 1.$$
Hence every pair $x \neq y$ maps to $(1,0)$. The operator is single-valued; hence $\mathcal{G}(I) = \{1\}$.

2. Choose any x, y,
$$Q_{cI+d}(x,y) = \frac{(cy+d)-(cx+d)}{\|y-x\|} = \frac{cy-cx}{\|y-x\|} = ce_{y-x}$$
$$s_{cI+d}(x,y) = \langle ce_{y-x}, e_{y-x}\rangle \pm i\sqrt{\|ce_{y-x}\|^2 - \langle ce_{y-x}, e_{y-x}\rangle^2} = c.$$
Hence $\mathcal{G}(cI+d) = \{c\}$.

3. Choose any $x \neq y$, and the rotation in $\mathbb{R}^2$ is $R_\theta = \begin{pmatrix} \cos\theta, \sin\theta \\ -\sin\theta, \cos\theta \end{pmatrix}$.
$$Q_{R_\theta}(x,y) = \frac{R_\theta y - R_\theta x}{\|y-x\|} = R_\theta e_{y-x}.$$
Using the definition of s_{R_θ} from equation (3b), it is clear that $s_{R_\theta} = (\cos\theta, \pm\sin\theta)$ and hence $\mathcal{G}(R_\theta) = \{(\cos\theta, \pm\sin\theta)\}$.

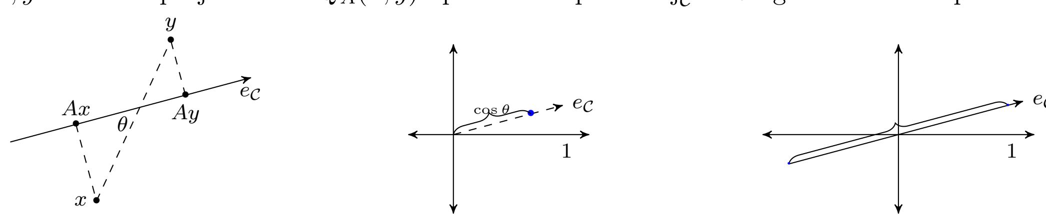
□

Example 2 (projection to line) Let $A : \mathbb{R}^2 \to \mathbb{R}^2$ be the projection onto a line $\mathcal{C}$, which we write as $A = \mathrm{Proj}_{\mathcal{C}}$ or $Ax = b + \langle x, e_{\mathcal{C}} \rangle e_{\mathcal{C}}$ where $e_{\mathcal{C}}$ is a unit vector in the direction of the line $\mathcal{C}$ and b is any point on the line. The quotient map of points $x \neq y$ is:

$$Q_A(x,y) = \frac{\mathrm{Proj}_{\mathcal{C}} y - \mathrm{Proj}_{\mathcal{C}} x}{\|y-x\|} = \frac{\langle y-x, e_{\mathcal{C}} \rangle}{\|y-x\|} e_{\mathcal{C}} = \langle e_{y-x}, e_{\mathcal{C}} \rangle e_{\mathcal{C}} = \cos\theta \cdot e_{\mathcal{C}}$$

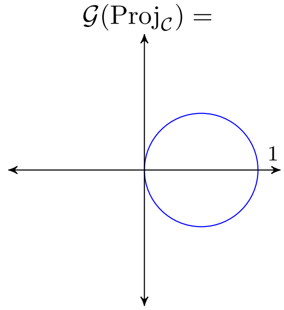
where $\theta = \angle(y-x, \mathcal{C})$ (the angle between vectors $y-x$ and $e_{\mathcal{C}}$).

x, y and their projections $\quad Q_A(x,y)$: quotient map of $\mathrm{Proj}_{\mathcal{C}}$ $\quad$ Range of values of quotient map



Let us now calculate the graphing map:

$\mathcal{G}(\mathrm{Proj}_{\mathcal{C}}) =$



$$\begin{aligned} s_A(x,y) &= \left(\langle Q_A(x,y), e_{y-x} \rangle, \sqrt{\|Q_A(x,y)\|^2 - \langle Q_A(x,y), e_{y-x} \rangle}^2 \right) \\ &= \left(\cos\theta \langle e_{\mathcal{C}}, e_{y-x} \rangle, i \sqrt{\|\cos\theta \cdot e_{\mathcal{C}}\|^2 - (\cos\theta \langle e_{\mathcal{C}}, e_{y-x} \rangle)^2} \right) \\ &= \left(\cos^2\theta, (\cos\theta)(\sin\theta) \right), \text{ where } \theta = \angle(y-x, \mathcal{C}) \\ &= \left(\frac{1}{2} + \frac{1}{2}\cos 2\theta, \frac{1}{2}\sin 2\theta \right) \end{aligned}$$

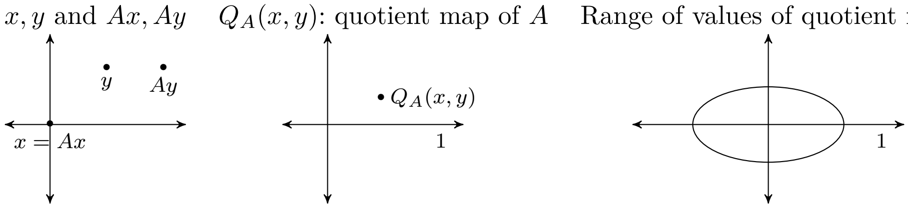
By plugging arbitrary $x \neq y$, we get all $\theta \in [0, 2\pi)$ and thus the graphing map traces out a circle of radius $1/2$ at $(1/2, 0)$. We have the SRG above since A is single-valued.

Example 3 (component-wise scaling) Let $A : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $A(u,v) = (2u, v)$. We need only consider the case where $x = (0,0)$ because A is a linear operator. If x is nonzero, we can just change coordinate systems to make x the origin. Let us use polar coordinates: $y = (u,v) = R(\cos\theta, \sin\theta)$.

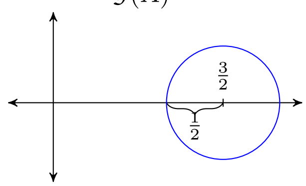
$$\begin{aligned} Q_A((0,0),(u,v)) &= \frac{1}{R}(2R\cos\theta, R\sin\theta) \\ &= (2\cos\theta, \sin\theta). \end{aligned}$$

Hence, the quotient map traces out an ellipse.

x, y and $Ax, Ay \quad Q_A(x,y)$: quotient map of A $\quad$ Range of values of quotient map



$\mathcal{G}(A) =$



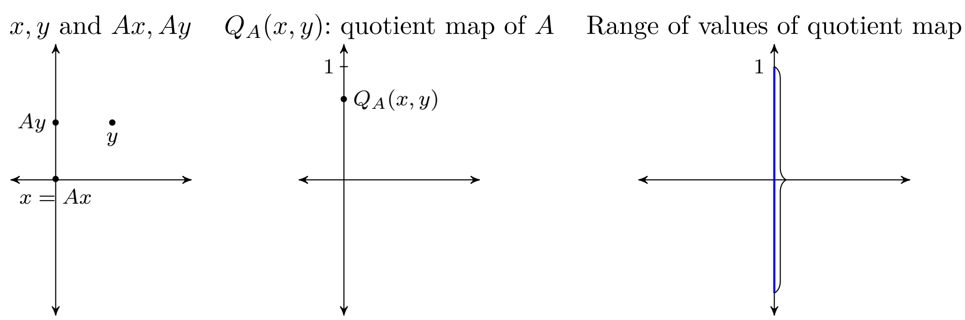
$$
\begin{aligned}
(s_A)_1 &= \langle (2\cos\theta, \sin\theta), (\cos\theta, \sin\theta) \rangle \\
&= 1 + \cos^2\theta = \frac{3}{2} + \frac{1}{2}\cos 2\theta \\
(s_A)_2 &= \pm\sqrt{\|(2\cos\theta, \sin\theta)\|^2 - (1+\cos^2\theta)^2} \\
&= \pm\sqrt{(1 + 3\cos^2\theta) - (1+\cos^2\theta)^2} \\
&= \pm\sqrt{\cos^2\theta - \cos^4\theta} \\
&= \pm\sin\theta\cos\theta = \pm\frac{1}{2}\sin 2\theta
\end{aligned}
$$

Hence the SRG is a circle of radius $1/2$ centered at $3/2$.

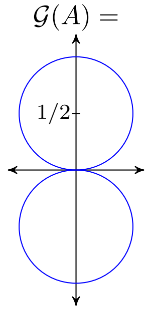
Example 4 (Diagonal matrix) Let $A(u,v) = (au, bv)$. A similar calculation yields that the SRG is a circle of radius $\frac{|b-a|}{2}$ centered around $\frac{b+a}{2}$.

Example 5 (Non-symmetric matrix) Consider the operator $A : \mathbb{R}^2 \to \mathbb{R}^2$ given by $A(u,v) = (0,u)$. As in Example 3, we set x to $(0,0)$ because A is linear. We use polar coordinates: $y = (u,v) = R(\cos\theta, \sin\theta)$.

$$Q_A((0,0),(u,v)) = (0, \cos\theta).$$

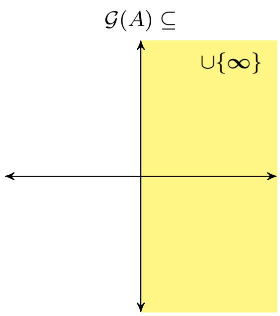


$\mathcal{G}(A) =$



$$
\begin{aligned}
(s_A)_1 &= \langle (0, \cos\theta), (\cos\theta, \sin\theta) \rangle \\
&= \cos\theta\sin\theta = \frac{1}{2}\sin 2\theta \\
(s_A)_2 &= \pm\sqrt{\|(0,\cos\theta)\|^2 - (\cos\theta\sin\theta)^2} \\
&= \pm\cos^2\theta = \pm\frac{1}{2}(1 + \cos 2\theta)
\end{aligned}
$$

Example 6 Let A be the gradient of the convex function $f(x_1, x_2) = x_1^4 + x_2^4$, that is, $A(x_1, x_2) = \nabla f(x_1, x_2) = (4x_1^3, 4x_2^3)$. Take any (x_1, x_2) and (y_1, y_2).



$$
\begin{aligned}
Q_A((x_1,x_2),(y_1,y_2)) &= \frac{(4x_1^3 - 4y_1^3, 4x_2^3 - 4y_2^3)}{\|(x_1-y_1, x_2-y_2)\|^2} \\
(s_A)_1 = Q_A((x_1,x_2),(y_1,y_2)) &\cdot ((x_1-y_1, x_2-y_2)) \\
&= \frac{4[(x_1-y_1)(x_1^3-y_1^3) + (x_2-y_2)(x_2^3-y_2^3)]}{\|(x_1-y_1,x_2-y_2)\|^2} \\
&= \frac{4[(x_1-y_1)^2(x_1^2+x_1y_1+y_1^2) + (x_2-y_2)^2(x_2^2+x_2y_2+y_2^2)]}{\|(x_1-y_1,x_2-y_2)\|^2} \geq 0
\end{aligned}
$$

That is, SRG of ∇f is contained right-hand plane. We'll see later that this implies that ∇f is a monotone operator.

4 Operator Properties, Transformations, and SRG

In this section we first study the correspondence between properties of operators and shapes of their SRGs. An operator's SRG can tell us much about the operator, such as whether the operator is averaged, cocoercive, Lipschitz, or strongly-monotone, etc. and by how much. For example, looking at an operator's SRG may allow us to deduce that the operator is $\frac{2}{3}$-averaged and $\frac{1}{2}$-Lipschitz. It also encodes properties like the real eigenvalues of linear operators, and whether an operator is injective. We demonstrate this correspondence by deriving strong if-and-only-if statements of the form: "An operator is X if and only if its SRG is Y". We then derive results to describe how transformations of an operator correspond to transformations of the SRG. Transformations such as translation, inversion, and adding the identity have simple effects on the SRG. These transformation rules will comprise the basic building blocks with which we will construct proofs in section 5.

Lemma 1 (Conversion Lemma) *Let $A : \mathbb{H} \rightrightarrows \mathbb{H}$ be an operator.*

1. Any inequality involving inner products, and norms of $Q_A(x,y)$ and e_{y-x} holds if and only if the corresponding inequality of inner products and norms holds for $s_A(x,y)$ and 1, respectively. For instance:

$$\|s_A(x,y) - c\|^2 \leq L^2 \iff \|Q_A(x,y) - ce_{y-x}\|^2 \leq L^2. \tag{5}$$

2. Any ***homogeneous*** *inequality involving inner products, and norms of $Ay - Ay$ and $y - x$ holds if and only if the corresponding inequality holds for inner products, and norms of $s_A(x,y)$ and 1 respectively. For instance*

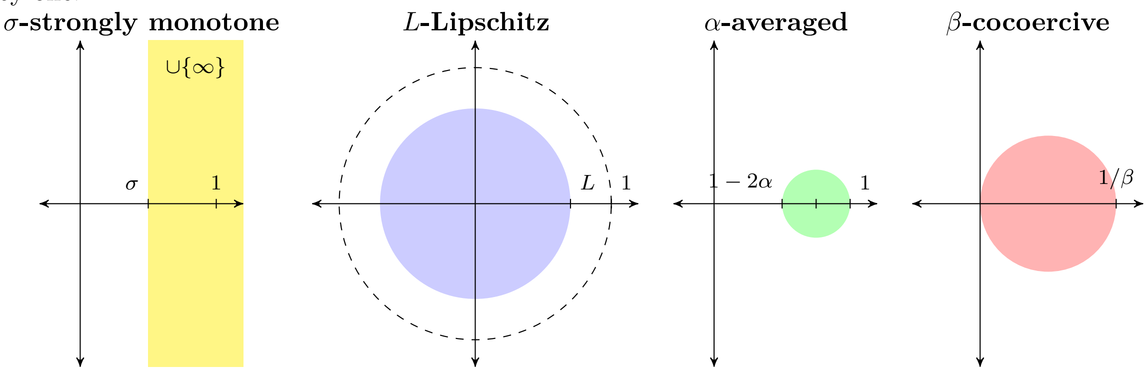
$$
\begin{aligned}
&\|Ay - Ax - (1-2\alpha)(y-x)\|^2 + \|Ay - Ax - (y-x)\|^2 \leq \alpha^2\|y-x\|^2 \\
\iff &\|s_A(x,y) - (1-2\alpha)1\|^2 \qquad + \|s_A(x,y) - 1\|^2 \qquad \leq \alpha^2.
\end{aligned}
$$

Proof 1. This is clear because $\|s_A(x,y)\| = \|Q_A(x,y)\|$, $\|e_{y-x}\| = 1$, and $\langle Q_A(x,y), e_{y-x}\rangle = \langle s_A(x,y), (1,0)\rangle$. Merely use the distributive law on all inner products and norms, and substitute the corresponding terms.

2. Simply divide the inequality involving $Ay - Ax$ and $y - x$ by the appropriate power of $\|y - x\|$ to make a corresponding inequality involving $Q_A(x,y)$ and e_{y-x} (in the example above, we divide by $\|y-x\|^2$). Then apply the first part of the lemma. □

4.1 Operator Properties and Graph Properties

Below is the summary of how operator properties correspond to SRG properties. We will prove these results one by one.



In the top row is an operator property. An operator has the property in the top row if and only if its SRG is contained in the closed region below. For example: An operator is 4-cocoercive if and only if its SRG is contained in the closed ball of radius $\frac{1}{8}$ at $(\frac{1}{8}, 0)$.

4.1.1 Monotone operators

Recall the definition of a monotone operator from Section 2.

Definition 1 An operator $A : \mathbb{H} \rightrightarrows \mathbb{H}$ is *monotone* if, for each selection $A' \in \mathcal{S}(A)$

$$\langle A'y - A'x, y - x\rangle \geq 0, \quad \forall x, y. \tag{6}$$

It is called *σ-strong monotone* for $\sigma \geq 0$ if, for each selection $A' \in \mathcal{S}(A)$,

$$\langle A'y - A'x, y - x\rangle \geq \sigma\|y - x\|^2, \quad \forall x, y. \tag{7}$$

A monotone operator can be thought of as a generalization of the idea of an increasing function. Monotone operators are important in optimization for many reasons: The most obvious being that a Gâteaux-differentiable function is convex if and only if its gradient is monotone [5, Proposition 17.10].

Example 7 Consider the gradient of the convex function $f(u, v) = \frac{1}{2}(u^2 + v^2 + v^4)$:

$$A(u, v) = \nabla f(u, v) = (u, v + 2v^3).$$

For any (p, q) and (u, v), A satisfies

$$\begin{aligned}
&\langle A(p, q) - A(u, v), (p, q) - (u, v)\rangle \\
=&\big\langle (p - u, (q - v) + 2(q^3 - v^3)), (p - u, q - v)\big\rangle \\
=&(p - u)^2 + (q - v)^2(1 + 2(q^2 + pv + v^2)) \\
\geq&(p - u)^2 + (q - v)^2 \\
=&\|(p - u, q - v)\|^2
\end{aligned}$$

Therefore, ∇f is 1-strongly monotone.

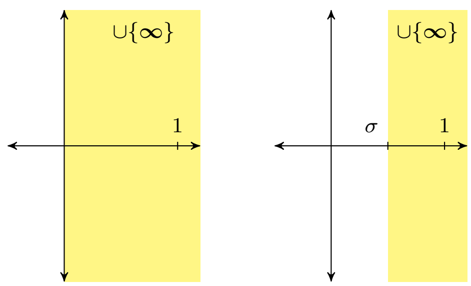
Proposition 1 *An operator A is **monotone** if and only if its SRG is contained in the right-hand plane* $([0,\infty)\times\mathbb{R})\cup\{\boldsymbol{\infty}\}$*. It is* σ-strongly monotone *if and only if its SRG is contained in* $([\sigma,\infty)\times\mathbb{R})\cup\{\boldsymbol{\infty}\}$*.*

Proof

$$\begin{aligned}
& A \text{ is } \sigma\text{-strongly monotone} \\
& \iff \langle A'y - A'x, y-x\rangle \geq \sigma\|y-x\|^2, \quad \forall x\neq y, \forall A'\in\mathcal{S}(A) \\
\text{Lem. 1 (Conversion Lemma)} & \iff \langle s_{A'}(x,y), 1\rangle \geq \sigma, \quad \text{"} \\
& \iff \operatorname{Re}(s_{A'}) \geq \sigma, \quad \text{"} \\
\text{(scaled relative graph definition)} & \iff \mathcal{G}(A) \subseteq \{(u,v)|u\geq\sigma\}\cup\{\boldsymbol{\infty}\}
\end{aligned}$$

□

Therefore, an operator is monotone and σ-strongly monotone, if its SRG is contained in the following regions respectively:



monotone **σ-strongly monotone**

Example 8 It can be shown that the closure of the SRG of the operator from Example 7 (i.e., $A(u,v) = (u, v+2v^3)$) fills the interior of $[1,\infty)\times\mathbb{R}$ (the diagram on the right, above) through the following steps. The SRG of a differentiable operator A contains the SRGs of the linear approximations of A at every point. The linear approximation of the operator $A(u,v) = (u, v+2v^3)$ at (u_0, v_0) is $(u, v(1+6v_0^2)) - (0, -6v_0^3)$. The SRG of this linear approximation is a circle between $(1,0)$ and $(1+6v_0^2, 0)$. The union of all these circles is the region $((1,\infty)\times\mathbb{R})\cup\{(1,0)\}$.

4.1.2 Lipschitz Operators

Recall the definition of a Lipschitz operator from Section 2.

Definition 2 An operator A is called ***L*-Lipschitz**, for $L\geq 0$, if it is single-valued and

$$\|Ay - Ax\| \leq L\|y-x\|, \quad \forall x\neq y. \tag{8}$$

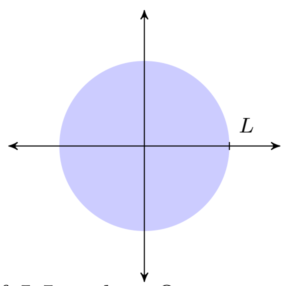
Lipschitz operators are important because they allow us to make simple estimates about how much an operator can change over a given distance. Convex functions in optimization algorithms are often assumed to have a Lipschitz gradients, which results in stronger convergence results.

Proposition 2 *An operator A is L-Lipschitz if and only if its SRG is contained in the closed ball of radius L at the origin: $\overline{B}(0,L)$.*

Proof

$$
\begin{aligned}
& && A \text{ is } L\text{-Lipschitz} \\
\text{Def. 2 (lipschitz op.)} & \iff && \frac{\|Ay - Ax\|}{\|y - x\|} \leq L; \qquad \forall x \neq y, \text{ and } A \text{ single-valued} \\
\text{Lem. 1 (Conversion Lemma)} & \iff && \|s_A(x, y)\| \leq L; \qquad \text{"} \\
\text{(SRG definition)} & \iff && \mathcal{G}(A) \subseteq \overline{B}(0, L).
\end{aligned}
$$

$\mathcal{G}(A) \subset$



SRG of L-Lipschitz Operator

Recall that $\mathcal{G}(A)$ does not include ∞ if and only if A is single-valued to complete the proof. □

4.1.3 Averaged Operators

Averagedness is a harder property to give an intuitive definition for. We can think of an averaged operator as a mixture of a nonexpansive operator and the identity. Averaged operators are important because, as discussed, the KM iteration of an averaged operator converges weakly to a fixed point of the operator. Recall the definition.

Definition 3 An operator A is α**-averaged**, $0 \leq \alpha \leq 1$, if there exists a nonexpansive operator R such that $A = (1 - \alpha)I + \alpha R$.

An averaged operator is $(1 - \alpha)$ parts identity and α parts a nonexpansive operator. The closer α is to 1, the more the operator will behave like a purely nonexpansive operator. The closer α is to 0, the more the operator will behave like the identity. For instance, the identity is monotone. When α becomes smaller than $\frac{1}{2}$, the operator becomes monotone just like the identity.

Example 9 Let $\mathcal{C}$ be a nonempty, closed, convex set. The projection onto $\mathcal{C}$ is 1/2-averaged. By standard results on projection to closed convex sets, the following relations hold:

$$
\begin{aligned}
0 &\geq \langle P_{\mathcal{C}} y - P_{\mathcal{C}} x, x - P_{\mathcal{C}} x \rangle, \\
0 &\geq \langle P_{\mathcal{C}} x - P_{\mathcal{C}} y, y - P_{\mathcal{C}} y \rangle,
\end{aligned}
$$

Summing the two inequalities yields:

$$
\begin{aligned}
\|P_{\mathcal{C}} x - P_{\mathcal{C}} y\|^2 &\leq \langle P_{\mathcal{C}} x - P_{\mathcal{C}} y, x - y \rangle, \\
\|(P_{\mathcal{C}} - \frac{1}{2}I)y - (P_{\mathcal{C}} - \frac{1}{2}I)x\|^2 &= \|P_{\mathcal{C}} y - P_{\mathcal{C}} x\|^2 - \langle P_{\mathcal{C}} x - P_{\mathcal{C}} y, x - y \rangle + \frac{1}{4}\|y - x\|^2 \leq \frac{1}{4}\|y - x\|^2.
\end{aligned}
$$

Hence $(2P_{\mathcal{C}} - I)$ is nonexpansive, and equivalently $P_{\mathcal{C}}$ is $\frac{1}{2}$-averaged.

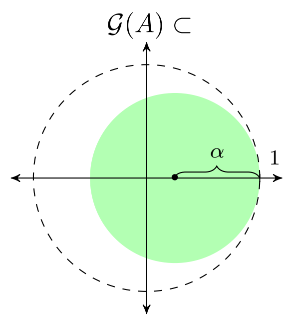
Proposition 3 *An operator A is α-averaged if and only if its SRG is contained in $\overline{B}(1-\alpha,\alpha)$.*

Proof

$$
\begin{aligned}
& && A \text{ is } \alpha\text{-averaged} \\
& \iff && A = (1-\alpha)I + \alpha R, \text{ with } R \text{ 1-Lipschitz} \\
& \iff && A - (1-\alpha)I \text{ is } \alpha\text{-Lipschitz.} \\
\text{(divide by } \|y-x\|^2) & \iff && \frac{\|(A-(1-\alpha)I)y - (A-(1-\alpha)I)x\|^2}{\|y-x\|^2} \le \alpha^2; \quad \forall x \neq y,\ A \text{ single valued} \\
\text{Lem. 1 (Conversion lemma)} & \iff && \|s_A(x,y) - 1 - \alpha\|^2 \le \alpha^2 \qquad \text{"} \\
\text{(SRG definition)} & \iff && \mathcal{G}(A) \subseteq \overline{B}(1-\alpha,\alpha).
\end{aligned}
$$

□

Therefore an operator is α-averaged if and only its SRG is contained in the following region:



SRG of α-Averaged Operator.

4.1.4 Cocoercive Operators

Cocoercive operators are important in monotone operator theory. Let f be convex and Fréchet differentiable. Then ∇f is Lipschitz (the most important class of differentiable functions) if and only if it is cocoercive. Co-coercive operators are also important in the convergence proof of the forward-backward algorithm, and for many other operator splitting algorithms. Recall the definition.

Definition 4 An operator A is **β-cocoercive** if it is single-valued and:

$$\beta\|Ay - Ax\|^2 \le \langle Ay - Ax, y - x\rangle, \quad \forall x \neq y. \tag{9}$$

Example 10 The operator $A(u,v) = (2u,v)$ from Example 3 is 1/2-cocoercive. Since A is linear, we may let $x = (0,0)$ without loss in generality.

$$\langle A(u,v),(u,v)\rangle = \langle (2u,v),(u,v)\rangle = 2u^2 + v^2 \ge \frac{1}{2}(4u^2+v^2) = \frac{1}{2}\|(2u,v)\|^2 = \frac{1}{2}\|A(u,v)\|^2.$$

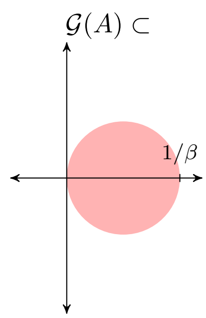
Proposition 4 *An operator A is β-cocoercive if and only if its graph is contained in $\overline{B}(\frac{1}{2\beta},\frac{1}{2\beta})$.*

Proof

$$
\begin{aligned}
& A \text{ is } \beta\text{-cocoercive} \\
& \iff \beta\|Ay - Ax\|^2 \le \langle Ay - Ax, y - x\rangle; \qquad \forall x \ne y, \text{ and single-valued.} \\
(\text{divide by } \|y-x\|^2) \; & \iff \beta\|Q_A(x,y)\|^2 \le \langle Q_A(x,y), e_{y-x}\rangle; \qquad \text{"} \\
& \iff \|Q_A(x,y)\|^2 - 2\frac{1}{2\beta}\langle Q_A(x,y), e_{y-x}\rangle \le 0; \text{"} \\
& \iff \|Q_A(x,y) - \frac{1}{2\beta}e_{y-x}\|^2 \le \left(\frac{1}{2\beta}\right)^2; \qquad \text{"} \\
\text{Lem. 1 (Conversion Lemma)} \; & \iff \|s_A(x,y) - \frac{1}{2\beta}\|^2 \le \left(\frac{1}{2\beta}\right)^2; \qquad \text{"} \\
(\text{SRG definition}) \; & \iff \mathcal{G}(A) \subseteq \overline{B}\left(\frac{1}{2\beta}, \frac{1}{2\beta}\right).
\end{aligned}
$$

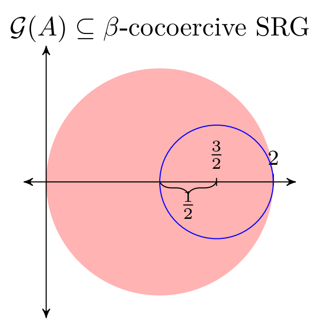
□

Therefore an operator is β-cocoercive if and only if its SRG is contained in the following region.



SRG of β-Cocoercive Operator.

Example 11 Let us demonstrate Proposition 4 with Example 3 again. We just calculated earlier that the operator A is 1/2-cocoercive. We also have from calculations in Example 3 that the SRG of the operator is a circle of radius 1/2 at $(3/2, 0)$. Hence the proposition holds in this case, as we can see in the diagram below.



4.1.5 Other properties

The SRG can tell us other properties about an operator.

Proposition 5 *1. An operator is* multi-valued *if and only if its SRG contains* ∞*.*
2. An operator is injective *if and only if its SRG does not contain* 0*.*
3. A linear operator has a real eigenvalue λ *if and only if* $(\lambda, 0)$ *is in its SRG.*

Proof 1. This is by definition.
2. The SRG of A contains 0 if and only if $Q_{A'}(x,y) = \frac{A'y - A'x}{\|y-x\|} = 0$ for some $A' \in \mathcal{S}(A)$ and $x \neq y$, or equivalently, $A'x = A'y$ for some $A' \in \mathcal{S}(A)$ and $x \neq y$ This is equivalent to A is not being injective.
3. Let (λ, v) be the an eigenvalue/vector pair for the linear operator A. Then, $Q_A(0,v) = \lambda v$ and hence $s_A(0,v) = \lambda$. Conversely, if $\lambda \in \mathcal{G}(A) \cap \mathbb{R}$ then there is an $x \in \mathbb{H}$ such that $\left\langle \frac{Ax}{\|x\|}, e_x \right\rangle = \lambda$, with Ax having no normal component with respect to e_x. Therefore, clearly x is a λ-eigenvector.

□

4.2 Operator Transformations and Graph Transformations

We now look at how transformations of an operator affect its SRG.

4.2.1 SRG-Invariant Transformations

We start with transformations that the graph is invariant under. First, however, we present a simple lemma to help determine when the SRG of an operator is unchanged.

Lemma 2 ***Invariant SRG Lemma:*** *Let* F *be a transformation we can apply to an operator (e.g. adding a constant:* $T \to T + \gamma$*, translating* $T \to T(\diamond - c)$*, etc.). Assume* F *is such that* T *is single-valued if and only if* $F(T)$ *is single-valued. If for this* F *there is a bijection* $\phi : \mathbb{H} \to \mathbb{H}$ *such that for every single-valued operator* $T' : \mathbb{H} \rightrightarrows \mathbb{H}$ *we have:*

$$s_{F(T')}(x,y) = s_{T'}(\phi(x), \phi(y)) \tag{10}$$

Then F *leaves the SRG unchanged.*

Proof Clearly $s_{F(T)}(x,y)$ is going to have the same range as $s_T(\phi(x), \phi(y))$ because ϕ is a bijection. Therefore the finite part of the SRG of any operator remains unchanged. The assumption we put on F means that $\infty \in \mathcal{G}(T)$ if and only if $\infty \in \mathcal{G}(F(T))$. □

Proposition 6 *The SRG is invariant under the following transformations:*

Addition of scalars:	$A + c$
Translation:	$A(x - c)$
Unitary change in coordinates:	$U \circ A \circ U^*$

Proof For all the above transformations, it is clear that A is single-valued if and only if $F(A)$ is single-valued.

1. Take $A' \in \mathcal{S}(A)$. We obtain:

$$Q_{A'+c}(x,y) = \frac{(A'y+c)-(A'x+c)}{\|y-x\|} = Q_{A'}(x,y), \quad \forall x \neq y.$$

   If $Q_{A'}(x,y)$ is unchanged by the addition of c, then so are $s_{A'}(x,y)$ and hence $\mathcal{G}(A)$ (using Lemma 2 with $\phi(x)=x$).
2. For any $A' \in \mathcal{S}(A)$, we have:

$$\begin{aligned} Q_{A'(\diamond-c)}(x,y) &= \frac{A'(y-c)-A'(x-c)}{\|(y-c)-(x-c)\|}, \forall x \neq y \\ e_{y-x} &= e_{(y-c)-(x-c)}, \qquad \text{"} \\ \text{Hence clearly: } s_{A'(\diamond-c)}(x,y) &= s'_A(x-c,y-c) \qquad \text{"} \end{aligned}$$

   Hence we can apply the invariant SRG Lemma with $\phi(x)=x-c$.
3. Take $A' \in \mathcal{S}(A)$.

$$\begin{aligned} Q_{U\circ A\circ U^*}(x,y) &= U \circ \frac{(A(U^*y)-A(U^*x))}{\|y-x\|}, \forall x,y \\ &= U \circ \frac{(A(U^*y)-A(U^*x))}{\|U^*(y-x)\|}, \forall x,y \\ &= UQ_A(U^*x,U^*y) \end{aligned}$$

   Hence

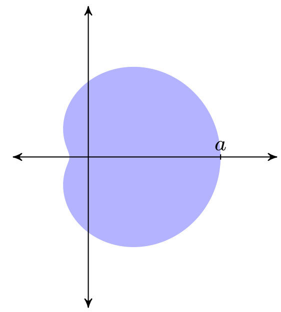
$$\begin{aligned} \|s_{U\circ A\circ U^*}(x,y)\| &= \|UQ_A(U^*x,U^*y)\| \\ &= \|Q_A(U^*x,U^*y)\| \\ &= \|s_A(U^*x,U^*y)\| \\ \mathrm{Re}(s_{U\circ A\circ U^*}(x,y)) &= UQ_A(U^*x,U^*y)\cdot e_{y-x} \\ &= Q_A(U^*x,U^*y)\cdot e_{U^*(y-x)} \\ &= s^1_A(U^*x,U^*y) \\ \text{Therefore } s_{U\circ A\circ U^*}(x,y) &= s_A(U^*x,U^*y) \end{aligned}$$

   Therefore, using lemma 2 again, with $\phi(x)=U^*x$ we have our result.

□

4.2.2 Transformation that Change the SRG

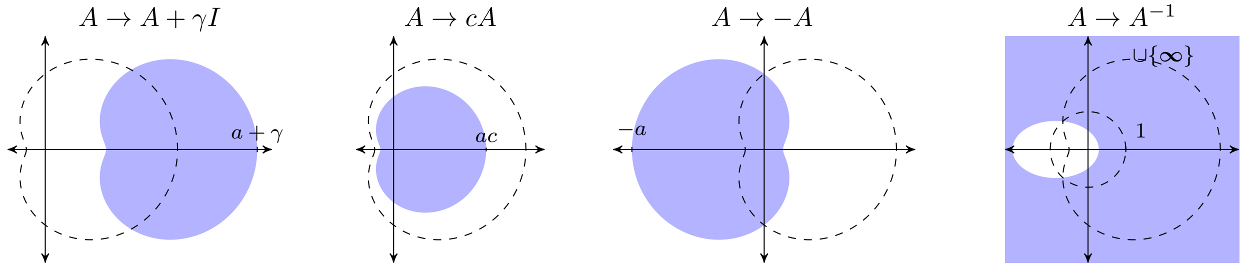
Other transformations have simple effects on the graph. We summarize these effects with a table, before proving them. We start with an operator $T : \mathbb{H} \to \mathbb{H}$ with the following SRG by way of example:



We then observe the effect that operator transformations have on the SRG. The top row describes the transformation of the operator. The second is graphical summary of what it does to the SRG. The bottom row is a more exact definition of what is occurring.

Operator Transformation vs. SRG Transformations:

$A \to A + \gamma I$ | $A \to cA$ | $A \to -A$ | $A \to A^{-1}$



Translate γ units right. $\mathcal{G}(A) \to \gamma + \mathcal{G}(A)$ | Scale by a factor of c. $\mathcal{G}(A) \to c\mathcal{G}(A)$ | Reflect about Im axis. $\mathcal{G}(A) \to -\mathcal{G}(A)$ | Invert through the unit circle. $\mathcal{G}(A) \to V(\mathcal{G}(A))$

We adopt the convention that shifting ∞ to the right, scaling ∞ by a factor c, or reflecting ∞ results in infinity again. Recall that V is the inversive map defined in Section 2, and that we take the convention that $1/0\infty$, and $1/\infty = 0$. We discuss V in more detail in the next chapter.

Proposition 7 *The following operator transformations on $A : \mathbb{H} \rightrightarrows \mathbb{H}$ have the following effect on A's SRG (for $c, \gamma \in \mathbb{R}$):*

Operator Transformation	***Effect on SRG***
$A \to A + \gamma I$	*Shift right γ units:* $\mathcal{G}(A) \to \gamma + \mathcal{G}(A)$
$A \to cA$	*Scale graph by factor of c for $c > 0$.* $\mathcal{G}(A) \to c\mathcal{G}(A)$
$A \to -A$	*Reflect about the vertical :* $\mathcal{G}(A) \to -\mathcal{G}(A)$
$A \to A^{-1}$	*Invert through the unit circle:* $\mathcal{G}(A) \to V(\mathcal{G}(A))$

Proof 1. A is multi-valued if and only if $A + \gamma I$ is multi-valued. If we take the convention that "moving ∞ to the right by γ units" results in ∞ again, we can move on to considering the "finite" part of the graph.

For $x \neq y$:

$$\begin{aligned} Q_{A+\gamma I}(x,y) &= \frac{(A+\gamma I)y - (A+\gamma I)x}{\|y-x\|} \\ &= \frac{Ay - Ax}{\|y-x\|} + \gamma \frac{y-x}{\|y-x\|} \\ &= Q_A(x,y) + \gamma e_{y-x} \end{aligned}$$

Using 2a, we have:

$$\begin{aligned} \operatorname{Re}(s_{A+\gamma I}) &= \operatorname{Re}(s_A) + \gamma \\ \operatorname{Im}(s_{A+\gamma I}) &= \pm\|(I - \operatorname{Proj}_{y-x}) Q_{A+\gamma I}(x,y)\| \\ &= \pm\|(I - \operatorname{Proj}_{y-x})(Q_A(x,y) + e_{y-x})\| \\ &= \pm\|(I - \operatorname{Proj}_{y-x}) Q_A(x,y)\| \\ &= \operatorname{Im}(s_A) \\ \text{Therefore: } s_{A+\gamma I}(x,y) &= s_A(x,y) + \gamma \end{aligned}$$

Thus we have a rightward shift of the graph by γ units.

2. For $c > 0$, A is multi-valued $\iff$ cA is multi-valued. If we take the convention that ∞ remains where it is if we expand the graph by a factor of c, we can move onto the finite part of the graph.

$$\begin{aligned} Q_{cA}(x,y) &= cQ_A(x,y) \\ \text{(Def. 3b) } s_{cA}(x,y) &= cs_A(x,y) \end{aligned}$$

So clearly we expand the graph by a factor of c because s_A is positively homogeneous in A.

3. $-A$ is multi-valued if and only if A is. For $x \neq y$, it can easily be shown that:

$$\begin{aligned} Q_{-A}(x,y) &= -Q_A(x,y) \\ \operatorname{Re}(s_{-A}(x,y)) &= -\operatorname{Re}(s_A(x,y)) \\ \operatorname{Im}(s_{-A}(x,y)) &= \operatorname{Im}(s_A(x,y)) \end{aligned}$$

Hence we reflect the SRG about the y-axis.

4. We deal with 0 and ∞ first. Recall:

$$\begin{aligned} \text{An operator } T \text{ is multi-valued} &\iff \infty \in \mathcal{G}(T) \\ \text{An operator } T \text{ is not injective} &\iff 0 \in \mathcal{G}(T) \end{aligned}$$

Applying this to A and A^{-1} (and recalling that a set-valued operator T is multi-valued if and only if its inverse T^{-1} is not injective) yields:

$$\begin{aligned} \infty \in \mathcal{G}(A) &\iff 0 \in \mathcal{G}\big(A^{-1}\big) \\ 0 \in \mathcal{G}(A) &\iff \infty \in \mathcal{G}\big(A^{-1}\big) \end{aligned}$$

Thus we have the result for the distinguished points in the SRG, 0 and ∞. We now consider the remainder of $\overline{\mathbb{R}^2}$. It is most convenient to use the definition of the graphing map from (3b):

$$\begin{aligned}
\mathcal{G}\big(A^{-1}\big)\backslash\{0,\infty\} &= \bigcup_{x\neq y, A'\in\mathcal{S}(A), A'x\neq A'y} \frac{\|y-x\|}{\|Ay-Ax\|}e^{\pm i\theta} \\
&= V\left(\bigcup_{x\neq y, A'\in\mathcal{S}(A), A'x\neq A'y} \frac{\|Ay-Ax\|}{\|y-x\|}e^{\pm i\theta}\right) \\
&= \bigcup_{z\in\mathcal{G}(A)\backslash\{0,\infty\}} V(z) = V(\mathcal{G}(A))
\end{aligned}$$

Thus we invert the SRG of A through the unit circle when we take A's inverse.

□

5 Proving Theorems with SRG

We now demonstrate how to prove theorems with SRG. The first theorems are not difficult to prove, and it may not be immediately clear why SRG should be used. SRG becomes more powerful later when we prove more complicated results: Especially those related to resolvents, convex functions, and rates of convergence.

Remark 1 (General Strategy) Say we have a theorem about a set-valued operator $T : \mathbb{H} \to 2^{\mathbb{H}}$ that we wish to prove:

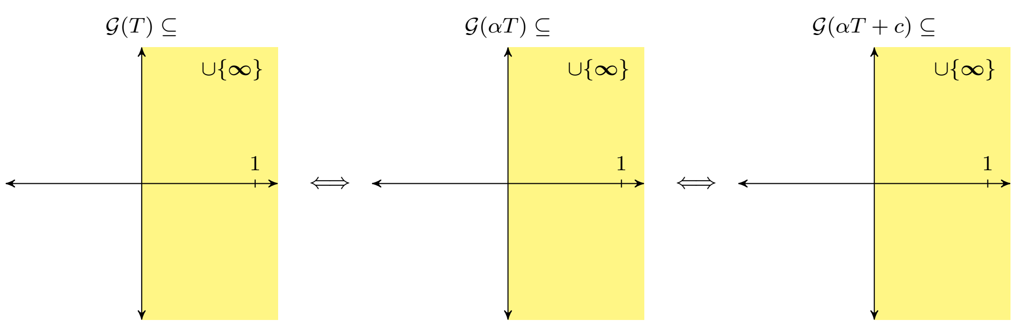
1. We take a statement about T as a starting point.
2. We convert this statement into a statement about T's SRG.
3. We use the transformation rules from Section 4 to chain together equivalent statements about the SRGs of a series of different operators.
4. We convert back into a statement about operators.

5.1 Monotone Operator Theorems

Proposition 8 *$T : \mathbb{H} \rightrightarrows \mathbb{H}$ is monotone if and only if $\alpha T + c$ is monotone for $\alpha > 0$, $c \in \mathbb{H}$.*

T is monotone: $\alpha T + c$ is monotone:

Proof



These diagrams can be transcribed as:

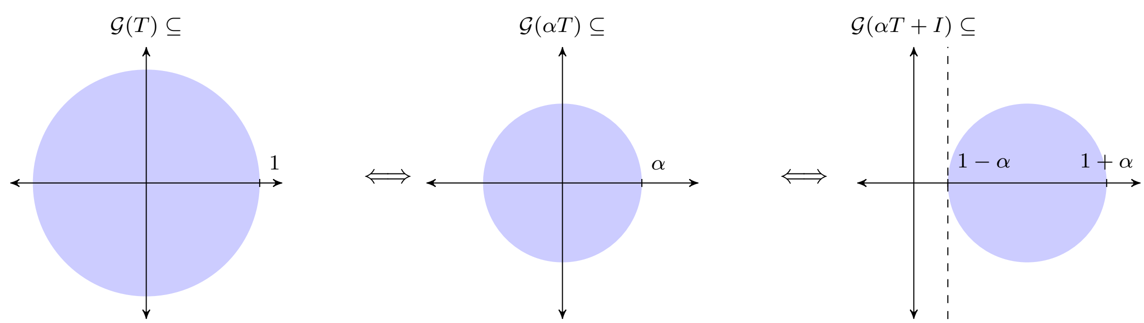
$$
\begin{aligned}
& T && \text{is monotone} \\
\text{(Prop. 1)} \iff & \mathcal{G}(T) && \subseteq \{(u,v)|u \geq 0\} \cup \{\infty\} \\
\text{(Prop. 7)} \iff & \mathcal{G}(\alpha T) && \subseteq \{(u,v)|u \geq 0\} \cup \{\infty\} \\
\text{(Prop. 6)} \iff & \mathcal{G}(\alpha T + c) && \subseteq \{(u,v)|u \geq 0\} \cup \{\infty\} \\
\text{(Prop. 1)} \iff & \alpha T + c && \text{is monotone.}
\end{aligned}
$$

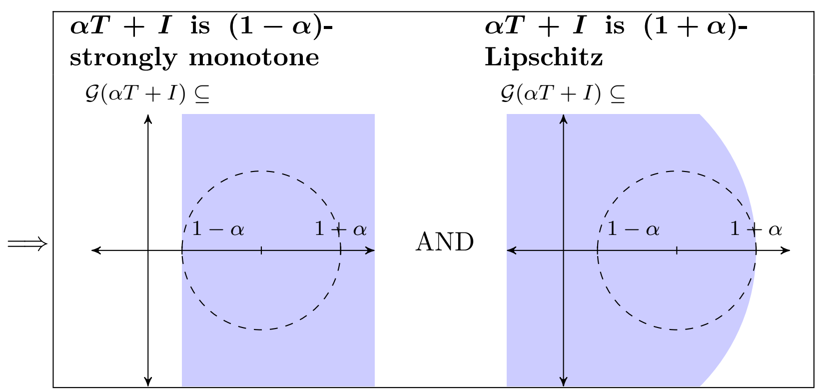
□

Proposition 9 *Let $T : \mathbb{H} \rightrightarrows \mathbb{H}$ be an nonexpansive operator, and $\alpha \in [0,1)$. Then $I + \alpha T$ is $(1-\alpha)$-strongly monotone, and $(1+\alpha)$-Lipschitz.*

Proof

T is nonexpansive





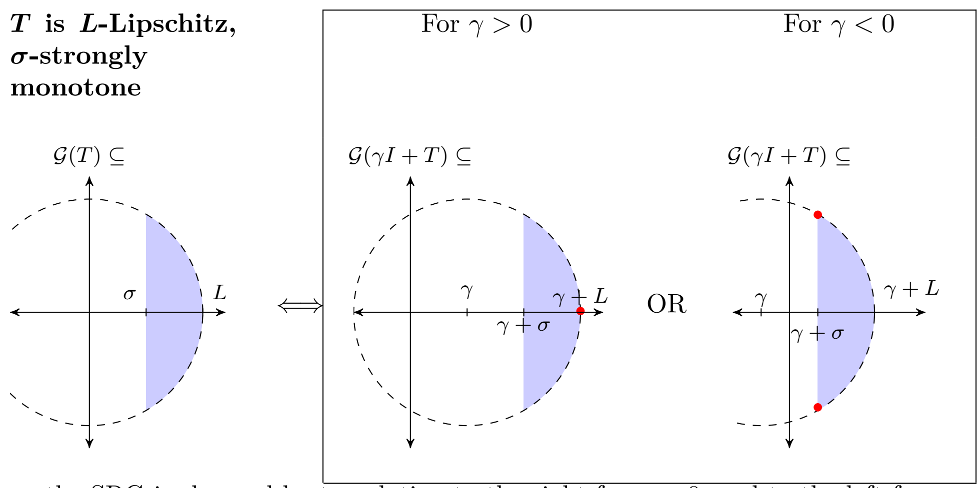
These diagrams can be transcribed as:

$$
\begin{aligned}
& T && \text{is nonexpansive} \\
\text{(Prop. 2)} \iff & \mathcal{G}(T) && \subseteq \overline{B}(0,1) \\
\text{(Prop. 7)} \iff & \mathcal{G}(\alpha T) && \subseteq \overline{B}(0,\alpha) \\
\text{(Prop. 7)} \iff & \mathcal{G}(I+\alpha T) && \subseteq \overline{B}(1,\alpha) \\
\text{(Props. 1, 2)} \implies & I+\alpha T && \text{is } \begin{cases} 1-\alpha\text{-strongly monotone} \\ 1+\alpha\text{-Lipschitz} \end{cases}
\end{aligned}
$$

□

Proposition 10 *Let $T : \mathbb{H} \rightrightarrows \mathbb{H}$ be an operator that is L-Lipschitz and σ-strongly monotone for $0 < \sigma < L$. Then $\gamma I + T$ is $(\gamma + L)$-Lipschitz for $\gamma > 0$ and $\sqrt{L^2 + \gamma^2 - 2\sigma\gamma}$-Lipschitz for $\gamma < 0$.*

Proof



As before, the SRG is changed by translating to the right for $\gamma > 0$, and to the left for $\gamma < 0$. To determine the Lipschitz constant, it is only necessary to find the extremal point of the SRG with simple geometry. These extremal points are denoted in red on both diagrams, and their distances from the origin are clearly $\gamma + L$ and $\sqrt{L^2 + \gamma^2 - 2\sigma\gamma}$ respectively. □

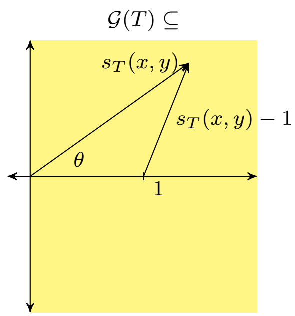
Proposition 11 *Let $T : \mathbb{H} \rightrightarrows \mathbb{H}$ be single-valued. Then T is monotone if and only if $I - T$ is* pseudocontractive*, that is:*

$$\|(I-T)y - (I-T)x\|^2 \leq \|y-x\|^2 + \|Ty - Tx\|^2, \forall x, y \tag{11}$$

Proof

$$
\begin{aligned}
& I - T \text{ pseudocontractive} \\
\iff & \|(Ty - Tx) - (y-x)\|^2 \leq \|y-x\|^2 + \|Ty-Tx\|^2; \quad \forall x, y \\
\text{(Conversion Lemma)} \iff & \|s_T(x,y) - 1\|^2 \leq \|1\|^2 + \|s_T(x,y)\|^2; \quad \text{"}
\end{aligned}
$$

Now consider the last statement above (we draw the situation below):



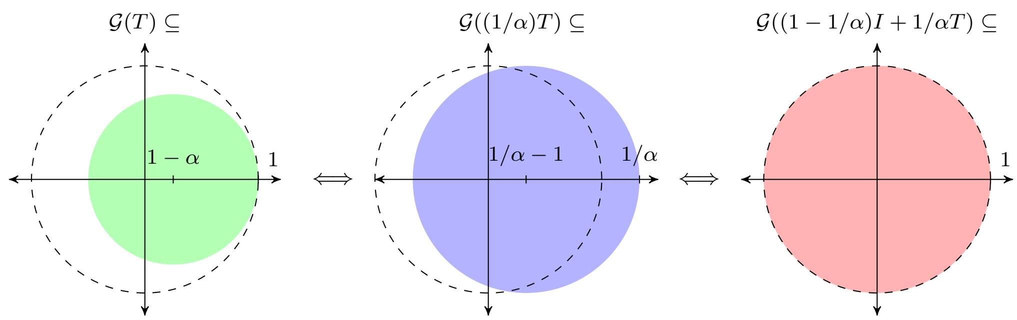
Notice $\|s_T(x,y)-e_1\|$, $\|e_1\|$, and $\|s_T(x,y)\|$ are side lengths of the triangle above. A standard geometry results states that this last inequality is equivalent to $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$. This describes exactly the right-hand plane. Therefore this is equivalent to T being monotone. □

5.2 Averaged and Cocoercive Operator Theorems

Proposition 12 *Let $T : \mathbb{H} \rightrightarrows \mathbb{H}$ be single-valued. Then the following are equivalent:*

1. *T is α-averaged.*
2. *$(1-\frac{1}{\alpha})I + \frac{1}{\alpha}T$ is nonexpansive.*

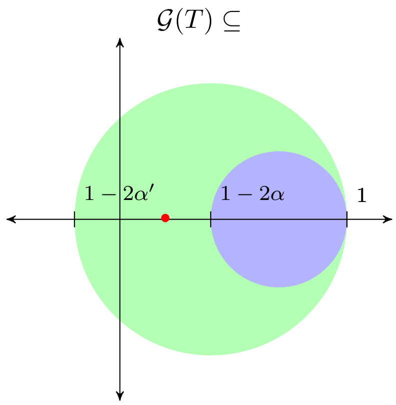
Proof **1 $\iff$ 2:**



□

Proposition 13 *Let $T : \mathbb{H} \rightrightarrows \mathbb{H}$ be single-valued. If T is α-averaged, then T is α'-averaged $\forall \alpha' > \alpha$. Strict inclusion applies: I.e. the class of α'-averaged operators is strictly larger than the class of α-averaged operators for $\alpha' > \alpha$.*

Proof



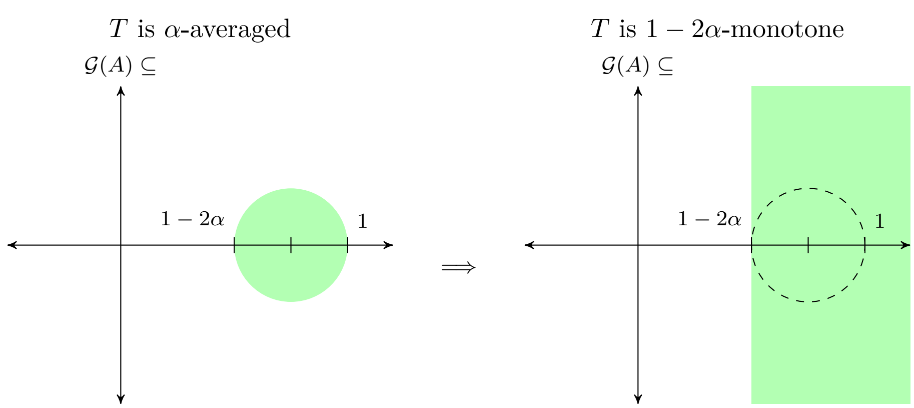
$$\begin{aligned} & T \text{ is } \alpha\text{-averaged} \\ \iff & \mathcal{G}(T) \subseteq \overline{B}(1-\alpha, \alpha) \\ \implies & \mathcal{G}(T) \subseteq \overline{B}(1-\alpha, \alpha'), \text{ since } \overline{B}(1-\alpha,\alpha) \subseteq \overline{B}(1-\alpha,\alpha') \\ \iff & T \ \alpha'\text{-averaged} \end{aligned}$$

To prove strict inclusion, we chose the obvious example $T = (1-2\hat{\alpha})I$ where $\alpha < \hat{\alpha} < \alpha'$. T is α'-averaged but not α-averaged. The SRG of T is a single point at $1-2\hat{\alpha}$ that can be seen to fall within the green region, but not the blue region. It is denoted with a red dot in the diagram.

□

Proposition 14 *Let $T : \mathbb{H} \rightrightarrows \mathbb{H}$ be α-averaged for $\alpha \in [0, \frac{1}{2}]$. Then T is monotone. In fact, for $\alpha \in [0, \frac{1}{2})$, T is $1-2\alpha$-strongly monotone.*

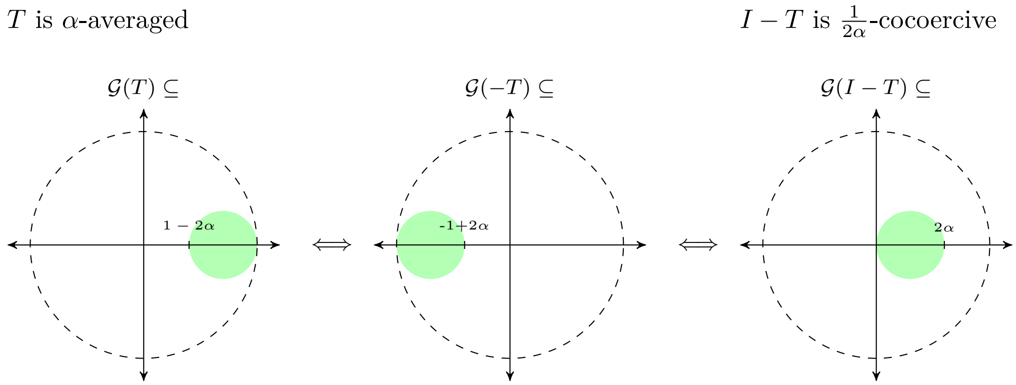
Proof



The proof is clear. □

Proposition 15 $T : \mathbb{H} \rightrightarrows \mathbb{H}$ *is α-averaged* $\iff$ $I - T$ *is $\frac{1}{2\alpha}$-cocoercive.*

Proof



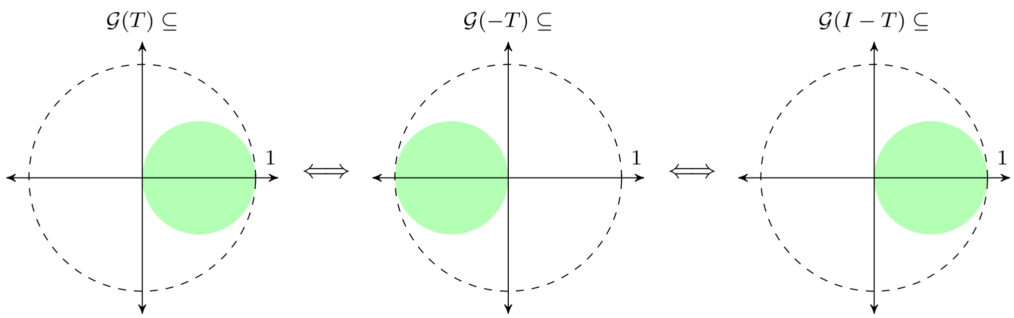
$$
\begin{aligned}
&T \text{ is } \alpha\text{-averaged}\\
\iff &\mathcal{G}(T) \subseteq \overline{B}(1-\alpha, \alpha)\\
\iff &\mathcal{G}(-T) \subseteq \overline{B}(\alpha-1, \alpha)\\
\iff &\mathcal{G}(I-T) \subseteq \overline{B}(\alpha, \alpha)\\
\iff &I + \alpha T \text{ is } \frac{1}{2\alpha}\text{-cocoercive}
\end{aligned}
$$

□

Proposition 16 *Let $T : \mathbb{H} \rightrightarrows \mathbb{H}$ be single-valued. Then the following are equivalent:*

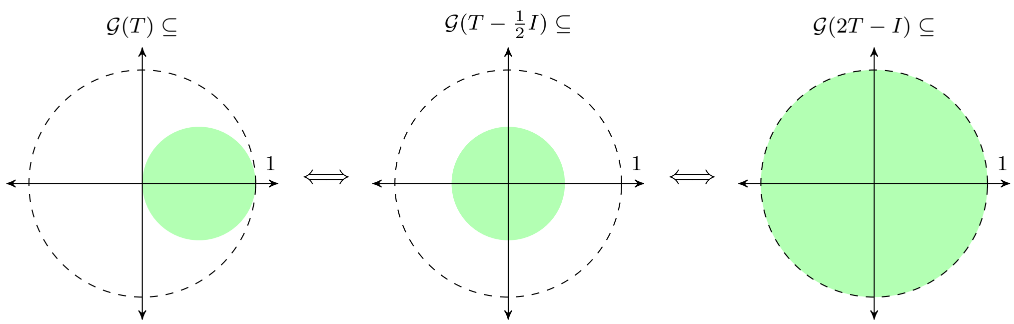
1. *T is FNE (i.e. 1-cocoercive)*
2. *$I - T$ is FNE*
3. *$2T - I$ is nonexpansive*
4. $\|Ty - Tx\|^2 + \|(I-T)y - (I-T)x\|^2 \le \|y - x\|^2$

1. T is FNE **2. $I - T$ is FNE**

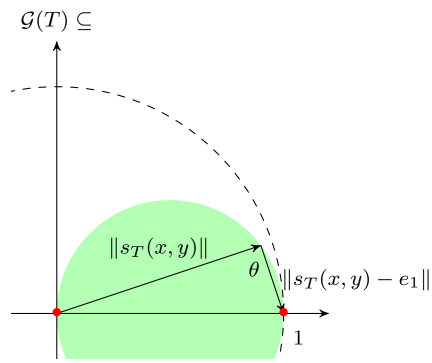


Proof

1. T is FNE **2. $2T - I$ is nonexpansive**



$$
\begin{aligned}
& \|Ty - Tx\|^2 + \|(I-T)y - (I-T)x\|^2 \leq \|y - x\|^2 \quad \forall x, y \\
\text{(Conversion lemma)} \iff & \|s_T(x,y)\|^2 + \|s_T(x,y) - e_1\|^2 \leq 1 \quad \text{"}
\end{aligned}
$$

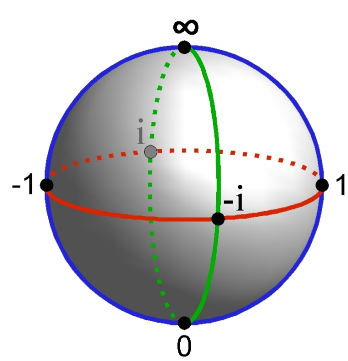


The first three equivalences are made via the standard transformations. Number 4 gives a statement about the SRG shown above (after applying the conversion lemma). Using the cosine rule, this statement is equivalent to $\theta \geq \pi/2$, which exactly describes the FNE circle. □

5.3 Operator Inversion

As discussed earlier, inverting an operator results in its SRG being inverted through the unit circle. Inversion through a circle brings generalized circles to generalized circles. Generalized circles are 1) circles or 2) lines union ∞. Lines are thought of as circles with infinite radius. We first define a notation for representing generalized circles.

Definition 5 Define $C(a,b,v)$ as the *closed* region of a generalized circle spanned by av and bv with axis of symmetry v, where $a, b \in \mathbb{R} \cup \{\infty\}$, $\|v\| = 1$.
If $a < b$, $C(a,b,v)$ represents the region within the circle.
$C(b,a,v)$ represents the region outside the circle (notice the swapping of the order).



We have to clarify what we mean by "closed" in this context. The extended 2D plane $\overline{\mathbb{R}^2}$ will be given the topology of a sphere (the Riemann Sphere). So for instance, the generalized circle $C(0,\infty,1)$ includes the point ∞.

Example 12 Let $c < a < 0 < b < \infty$, and $v = (1/\sqrt{2}, /\sqrt{2})$.

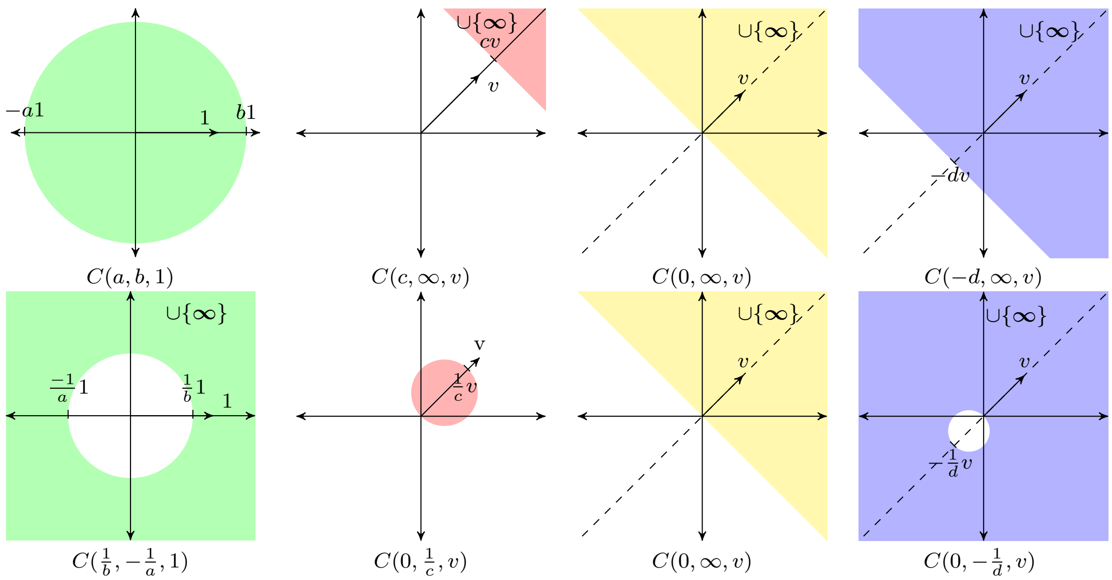
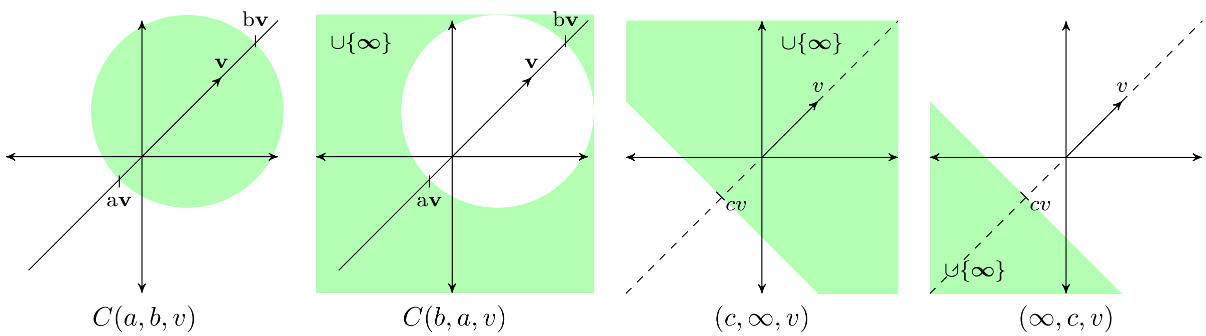


Fig. 1 Generalized Circle Inversion Pairs



Using this notation, we have a very simple formula for the action of V on generalized circles.

Lemma 3 *If we invert the generalized circle $C(a,b,v)$ through the unit circle, we obtain generalized circle $C(1/b,1/a,v)$ (noticing again the swapping of the order).*

We give some simple examples of using this lemma in Figure 13.

Example 13 Let $a,b,c,d > 0$, and $v = (1/\sqrt{2},/\sqrt{2})$. In rows 1 and 3, we present generalized circles $C(a,b,1)$, $C(c,\infty,v)$, $C(0,\infty,v)$, $C(-d,\infty,v)$. The corresponding entry just below (with matching color) is the same generalized circle inverted through the unit circle. That is, rows 2 and 4 contain: $C(1/b,-1/a,1)$, $C(0,1/c,v)$, $C(0,\infty,v)$, $C(0,-1/d,v)$ respectively.

5.4 Resolvent and Yosida Approximation

The resolvent and Yosida approximation of a monotone operato often appear in algorithms (for instance, the forward-backward, Douglass-Rachford, and ADMM algorithms).

Definition 6 Let $A : \mathbb{H} \rightrightarrows \mathbb{H}$ be an operator, and $\gamma > 0$ be a real number. A's **resolvent** is defined as:

$$J_A = (I + A)^{-1} \tag{12}$$

A's **reflected resolvent** is defined as:

$$R_{\gamma A} = 2J_{\gamma A} - I \tag{13}$$

A's **Yosida approximation** of index γ is defined as:

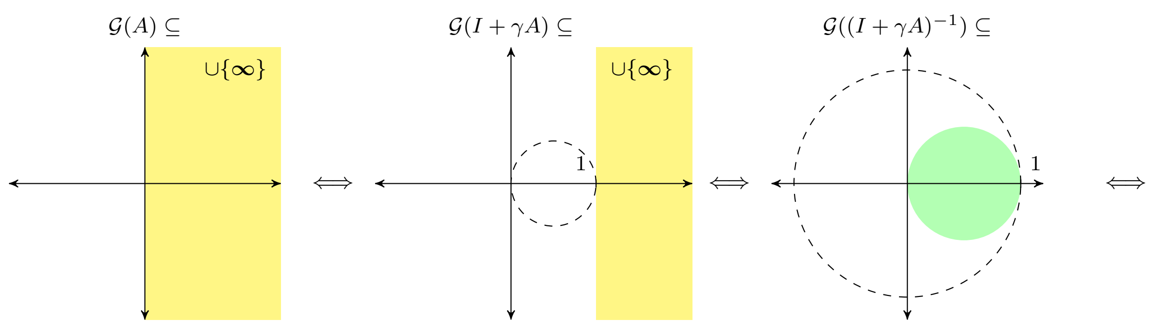
$$^{\gamma}A = \frac{1}{\gamma}(I - J_{\gamma A}) \tag{14}$$

Proposition 17 *Let $A : \mathbb{H} \rightrightarrows \mathbb{H}$ be an operator. The following are equivalent:*
1. A is monotone.
2. $J_{\gamma A}$ is firmly nonexpansive.
3. $^{\gamma}A$ is γ-cocoercive.

Proof We have:

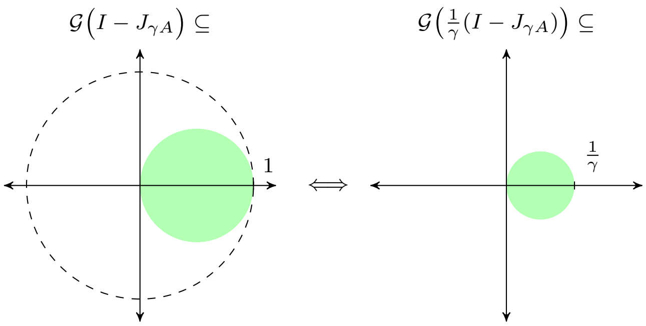
1: A is monotone

2: $J_{\gamma A} = (I + \gamma A)^{-1}$ is FNE



$I - J_{\gamma A}$ is FNE

3: ${}^{\gamma}A = \frac{1}{\gamma}(I - J_{\gamma A})$ is γ-cocoercive

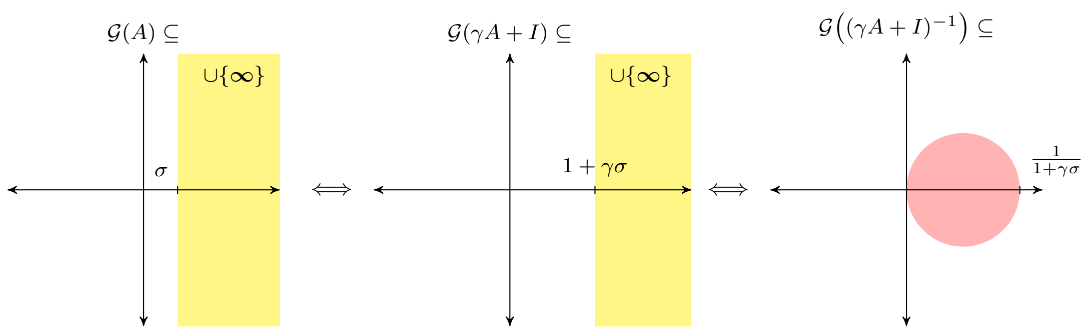


□

Proposition 18 *Let $A : \mathbb{H} \rightrightarrows \mathbb{H}$ be a monotone operator. A is σ- strongly monotone if and only if $J_{\gamma A}$ is $1 + \gamma\sigma$-cocoercive.*

Proof
A is σ-strongly monotone

$(1 + \gamma A)^{-1}$ is $1 + \gamma\sigma$-cocoercive.

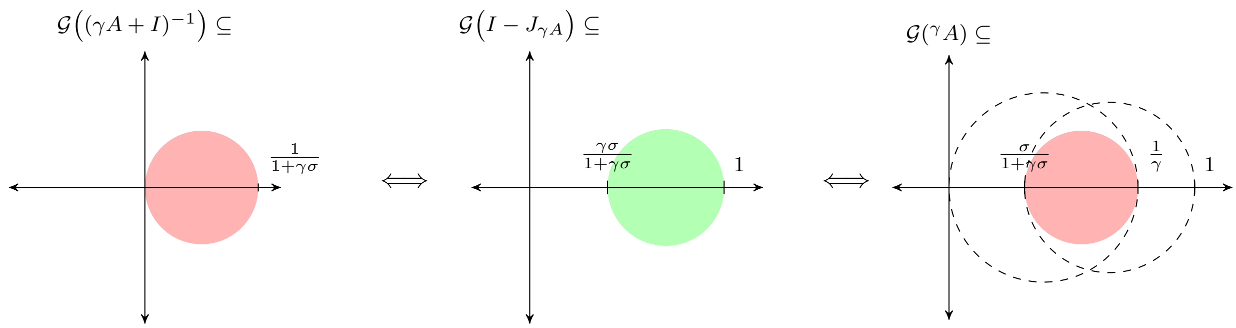


□

5.5 Proof of Theorem 1

Proposition 19 *Let $A : \mathbb{H} \to 2^{\mathbb{H}}$. If A is σ-strongly monotone, we have:*
1. ${}^{\gamma}A$ is still γ-cocoercive. I.e. ${}^{\gamma}A$ is not γ'-cocoercive for $\gamma' > \gamma$ in general. A being strongly monotone does not change ${}^{\gamma}A$'s cocoercivity.
2. ${}^{\gamma}A$ is also $\frac{1}{2}\left(1 - \frac{\sigma}{1+\sigma\gamma}\right)$-averaged.
3. ${}^{\gamma}A$ is $\frac{\sigma}{1+\gamma\sigma}$-strongly monotone.

Proof Let A be σ-strongly monotone. From the previous proposition we have the first SRG shown below.



${}^{\gamma}A$ is clearly γ cocoercive. Simple algebra shows that the SRG of ${}^{\gamma}A$ is contained in:

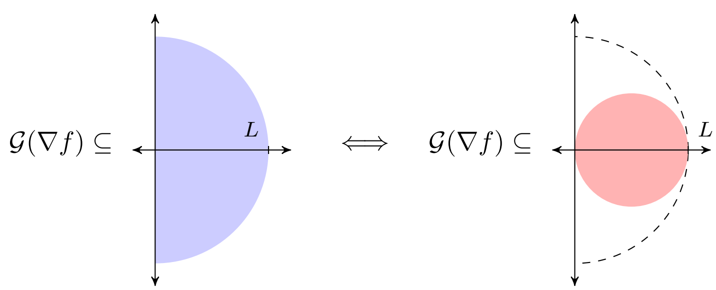
$$\overline{B}\left(1 - \frac{1}{2}\left(1 - \frac{\sigma}{1+\sigma\gamma}\right), \frac{1}{2}\left(1 - \frac{\sigma}{1+\sigma\gamma}\right)\right)$$

Therefore, ${}^{\gamma}A$ is $\frac{1}{2}\left(1 - \frac{\sigma}{1+\sigma\gamma}\right)$-averaged. It can also be readily seen that ${}^{\gamma}A$ is $\frac{\sigma}{1+\gamma\sigma}$-strongly monotone, since ${}^{\gamma}A$'s SRG is contained in $[\sigma/(1+\gamma\sigma), \infty) \times \mathbb{R}$. □

6 The Baillon–Haddad Theorem

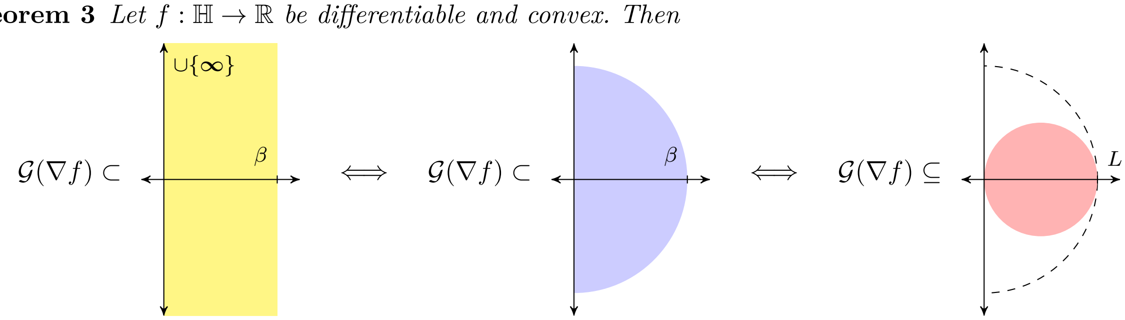
The classical Baillon–Haddad theorem can be expresse elegantly with the SRG.

Theorem 2 (Baillon–Haddad [2]) *Let $f : \mathbb{H} \to \mathbb{R}$ be differentiable and convex. Then*



In other words, ∇f being Lipschitz implies ∇f is $(1/L)$-cocoercive. This result was strengthened by Bauschke and Combettes

Theorem 3 *Let $f : \mathbb{H} \to \mathbb{R}$ be differentiable and convex. Then*

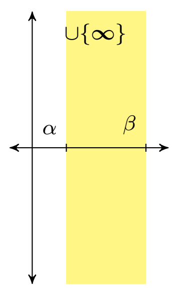


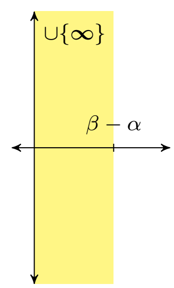
This extension was presented by Bauschke and Combettes [4] although their presentation did not use the notion of SRG.

From this it immediately follows:

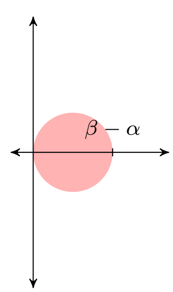
Corollary 1 *Let f be differentiable, and let $-\infty < \alpha < \beta < \infty$. Then $\frac{1}{2}\alpha\|x\|^2 + f$ and $\frac{1}{2}\beta\|x\|^2 - f$ are convex if and only if $\mathcal{G}(\nabla f) \subset \bar{B}(\frac{\beta+\alpha}{2}, \frac{\beta-\alpha}{2})$. In particular, f is α-strongly convex and β-smooth if and only if $\mathcal{G}(\nabla f) \subset \bar{B}(\frac{\beta+\alpha}{2}, \frac{\beta-\alpha}{2})$.*

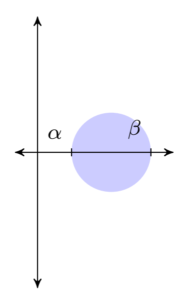
Proof Simply apply Theorem 3 to $f - \frac{1}{2}\alpha\|x\|^2$. The SRG of the gradient of an α-strongly convex and β-smooth function f, $\mathcal{G}(\nabla f)$, is shown as below. Moreover, we shift $\mathcal{G}(\nabla f)$ to the left α units to obtain $\mathcal{G}(\nabla f - \alpha I)$.

$\mathcal{G}(\nabla f) \subset$ 

$\iff$ $\mathcal{G}(\nabla f - \alpha I) \subset$ 

By Theorem 3,

$\iff$ $\mathcal{G}(\nabla f - \alpha I) \subset$ 

$\iff$ $\mathcal{G}(\nabla f) \subset$ 

□

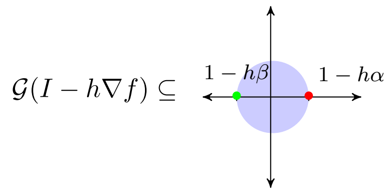
7 Convergence Rates

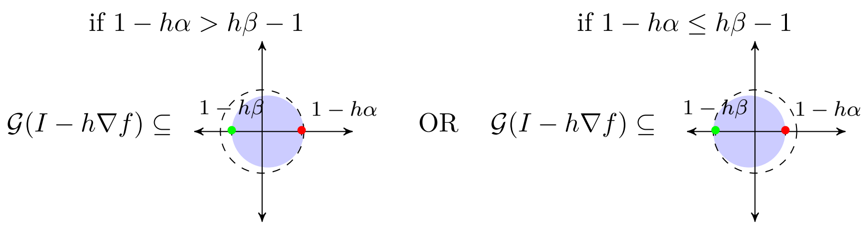
We now use SRG to derive some convergence rates of algorithms.

7.1 Gradient Method

Proposition 20 *Let f be α-strongly convex with a β-Lipschitz gradient. Say that it has a minimizer x^*. The gradient descent algorithm $x^{k+1} = x^k - h\nabla f(x^k) = (I - h\nabla f)(x^k)$ has convergence rate $\mathcal{O}(\max(1 - h\alpha, h\beta - 1)^k)$. The speed of convergence is maximized when $h = \frac{2}{\alpha+\beta}$, which corresponds to a rate of $1 - \frac{2\alpha}{\alpha+\beta}$.*

Proof From Corollary 1, we have



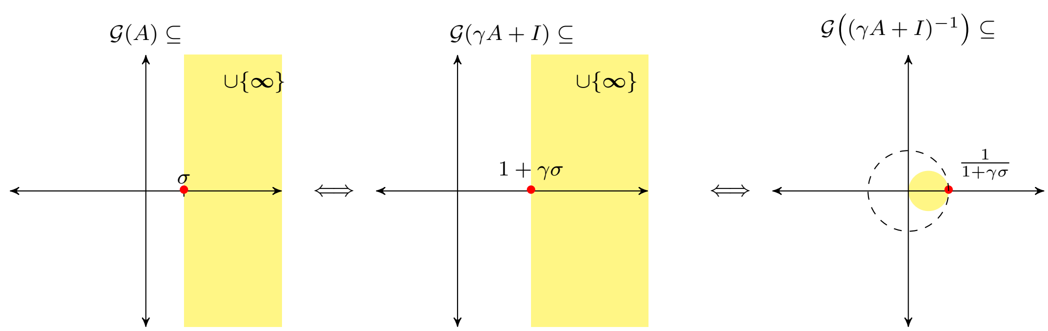


Therefore the convergence rate is $\max_h\{(1 - h\alpha, h\beta - 1)\}$. □

7.2 Proximal Algorithms

Proposition 21 *Let $A : \mathbb{H} \rightrightarrows \mathbb{H}$ be a σ-strongly maximally monotone. The proximal-point algorithm $x^{k+1} = J_{\gamma A} x^k$ converges to a zero of A with rate $\mathcal{O}((1 + \gamma\sigma)^{-k})$.*

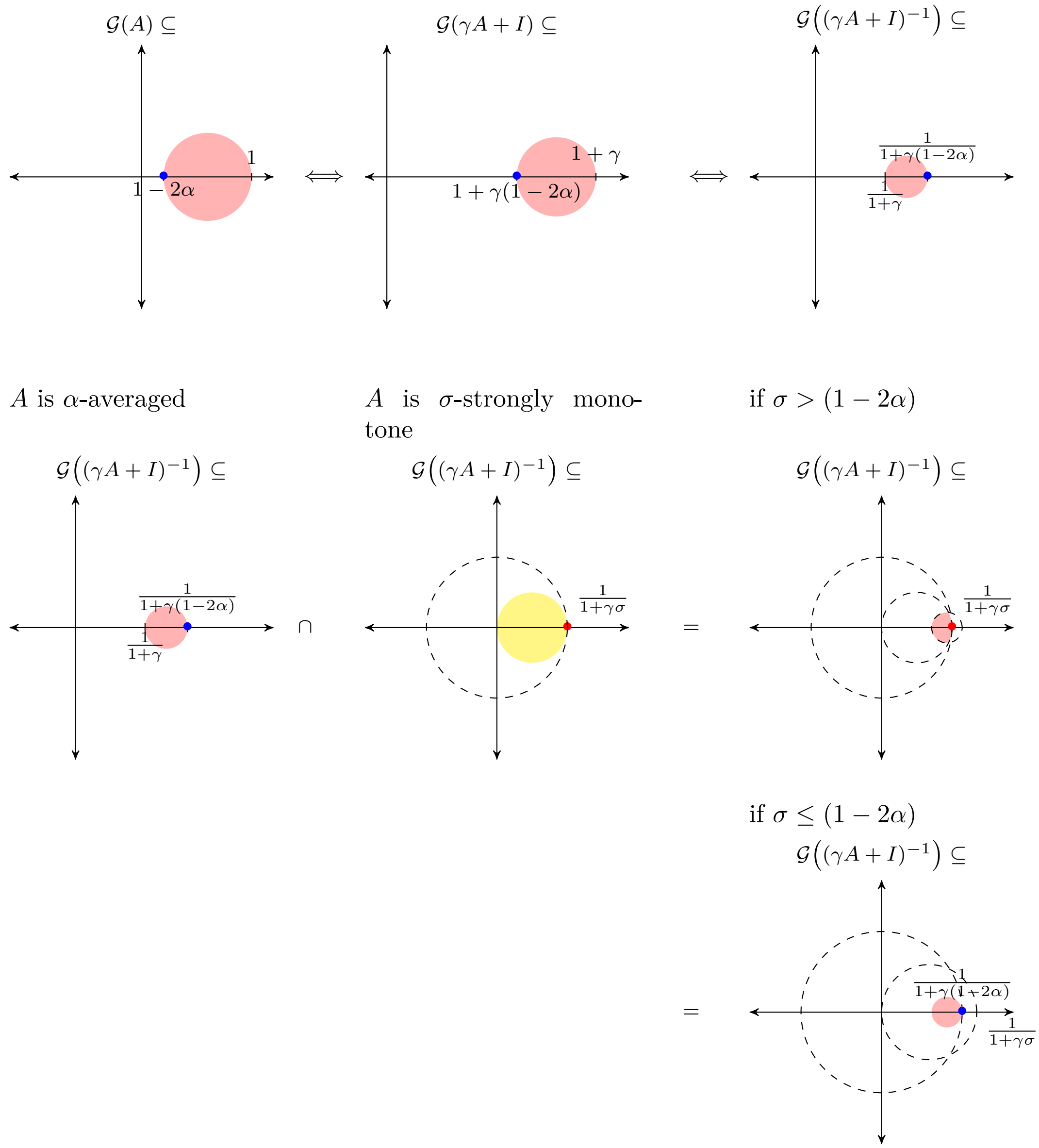
Proof Since A is σ-strongly monotone



The final diagram implies that $(I + \gamma A)^{-1}$ is $(1 + \gamma\sigma)^{-1}$-Lipschitz. □

Proposition 22 *Let $A : \mathbb{H} \rightrightarrows \mathbb{H}$ be a σ-strongly maximally monotone for $\sigma \leq 1$, and α-averaged. Then proximal-point algorithm $x^{k+1} = J_{\gamma A} x^k$ converges to a zero of A with rate $\mathcal{O}((1 + \gamma \max\{\sigma, 1 - 2\alpha\})^{-k})$.*

Proof Since A is α-averaged



□

Definition 7 Let $A : \mathbb{H} \rightrightarrows \mathbb{H}$ be maximally monotone. The **over-relaxed proximal point algorithm** is defined via:

$$x^{k+1} = (I(1-\lambda) + \lambda J_{\gamma A})x^k$$

for an averaging parameter λ.

Proposition 23 *Let $A : \mathbb{H} \rightrightarrows \mathbb{H}$ be a σ-strongly maximally monotone. Then the over-relaxed proximal-point algorithm has rates of convergence as follows:*

$$\|x^k - x^*\| \leq \|x^0 - x^*\| \left(\max \left(1 - \frac{\lambda\gamma\sigma}{1+\gamma\sigma},\ |1-\lambda| \right) \right)^k \text{ for } \lambda \geq 1$$

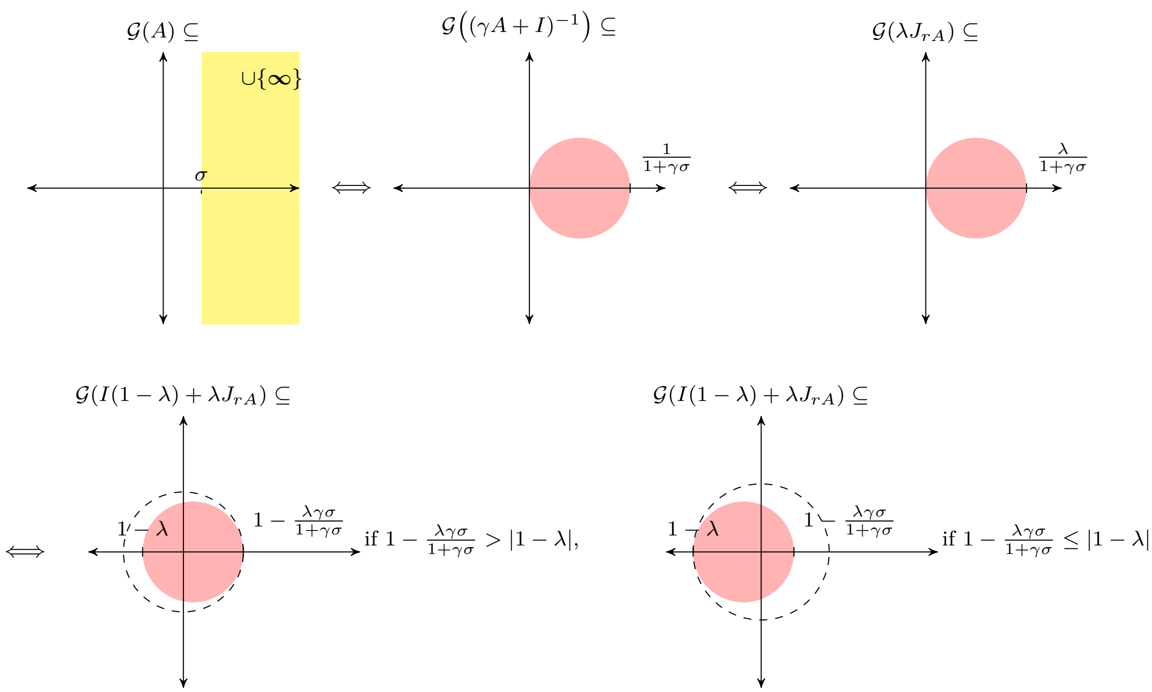
This rate is maximized when:

$$\lambda = 1 + \frac{1}{1+2\gamma\sigma}$$

leading to:

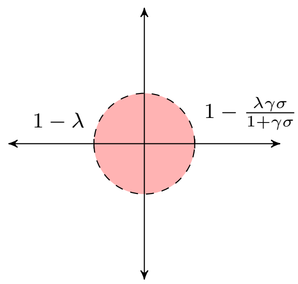
$$\|x^k - x^*\| \leq \|x^0 - x^*\| \left(\frac{1}{1+2\gamma\sigma} \right)^k \text{ for } \lambda \leq 1$$

Proof



Hence it can clearly be seen that the convergence rate is $\max_{\lambda}\{\lambda - 1, 1 - \frac{\gamma\sigma}{1+\gamma\sigma}\lambda\}$. The rate is maximized when this expression reaches its minimizer, which corresponds to the following:

$\mathcal{G}\Big((\gamma A + I)^{-1}\Big) \subseteq$



$$|1-\lambda| = 1 - \frac{\lambda\gamma\sigma}{1+\gamma\sigma}, \lambda \geq 1,$$
$$\lambda = 1 + \frac{1}{1+2\gamma\sigma}$$

It can easily be verified that this value of λ corresponds to rate $\frac{1}{1+2\gamma\sigma}$. $\square$

References

1. Ablowitz, M.J., Fokas, A.S.: Complex Variables: Introduction and Applications, 2nd edn. Cambridge University Press (2003)
2. Baillon, J.B., Haddad, G.: Quelques propriétés des opérateurs angle-bornés etn-cycliquement monotones **26**(2), 137–150. URL https://doi.org/10.1007/BF03007664
3. Banach, S.: Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales. Fundamenta Mathematicae **3**(1), 133–181 (1922)
4. Bauschke, H.H., Combettes, P.L.: The Baillon-Haddad theorem revisited URL http://arxiv.org/abs/0906.0807
5. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer Science & Business Media
6. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces, 2nd edn. Springer-Verlag (2017)
7. Beck, A.: First-Order Methods in Optimization. Society for Industrial and Applied Mathematics (2017)
8. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press (2004)
9. Eckstein, J.: Splitting methods for monotone operators with applications to parallel optimization. Ph.D. thesis, MIT (1989)
10. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Mathematical Programming **55**(1–3) (1992)
11. Giselsson, P.: Lunds universitet, lecture notes: Large-scale convex optimization (2015). URL: http://www.control.lth.se/education/doctorate-program/large-scale-convex-optimization/. Last visited on 2018/12/01
12. Giselsson, P., Boyd, S.: Metric selection in Douglas-Rachford splitting and ADMM URL http://arxiv.org/abs/1410.8479
13. Hiriart-Urruty, J.B., Lemaréchal, C.: Convex Analysis and Minimization Algorithms, vol. 2. Springer-Verlag (1993)
14. Krasnosel'skii, M.A.: Two remarks on the method of successive approximations. Uspekhi Matematicheskikh Nauk **10**(1), 123–127 (1955)
15. Mann, W.R.: Mean value methods in iteration. Proceedings of the American Mathematical Society **4**(3), 506–510 (1953)
16. Morley, F., Morley, F.V.: Inversive Geometry. G. Bell and Sons (1933)
17. Pedoe, D.: A Course Geometry for Colleges and Universities. Cambridge University Press (1970)
18. Ryu, E.K., Boyd, S.: Primer on monotone operator methods. Appl. Comput. Math. **15**, 3–43 (2016)
19. Wentworth, G., Smith, D.E.: Plane and Solid Geometry. Ginn and Company (1913)