Scaled Relative Graph:
Nonexpansive operators via 2D Euclidean Geometry

Ernest K. Ryu (류경석)

Department of Mathematical Sciences
Seoul National University

Joint work with Robert Hannah and Wotao Yin

Variational Analysis and Optimisation Seminar

November 25, 2020

Motivation

Fixed-point iterations are widely used in applied and computational mathematics.

Convergence of fixed-point iterations is usually established analytically with inequalities. Such proofs are often unintuitive.

We introduce an alternate geometric approach based on elementary 2D geometry. These proofs are visual and intuitive.

Talk based on [1].

[1] **Ryu**, Hannah, Yin, Scaled Relative Graph: Nonexpansive Operators via 2D Euclidean Geometry, under revision, 2019.

A sample result

Fact

Assume f is μ-strongly convex and L-smooth. Then

$$x^{k+1} = x^k - \alpha \nabla f(x^k)$$

converge exponentially to the minimizer $x^\star$ with rate

$$\|x^k - x^\star\| \leq \left(\max\{|1-\alpha\mu|, |1-\alpha L|\}\right)^k \|x^0 - x^\star\|.$$

A sample result

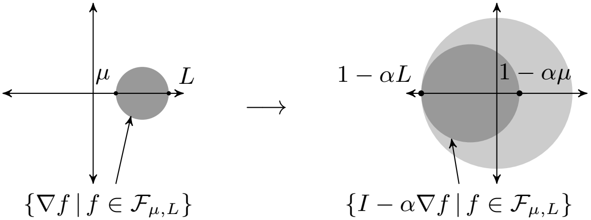
Fact

Assume f is μ-strongly convex and L-smooth. Then

$$x^{k+1} = x^k - \alpha \nabla f(x^k)$$

converge exponentially to the minimizer $x^\star$ with rate

$$\|x^k - x^\star\| \leq \left(\max\{|1-\alpha\mu|, |1-\alpha L|\}\right)^k \|x^0 - x^\star\|.$$



We make this illustration a rigorous proof.

Outline

Background and preliminaries

Scaled relative graph

Operator and SRG transformation

Fixed-point terations

Using a fixed-point iteration:

1. Find an operator $T : \mathcal{H} \to \mathcal{H}$ such that if $x^\star = T(x^\star)$ then $x^\star$ is a solution to the problem at hand.
2. Perform the **fixed-point iteration**

$$x^{k+1} = T(x^k).$$

Convergence via operator properties: nonexpansive

$T : \mathcal{H} \to \mathcal{H}$ is **nonexpansive** if

$$\|T(x) - T(y)\| \leq \|x - y\| \quad \forall\, x, y \in \mathcal{H}.$$

Fixed-point iterations with nonexpansive operators need not converge.
(E.g. $T(x) = -x$.)

Convergence via operator properties: contractive

$T : \mathcal{H} \to \mathcal{H}$ is **contractive** if

$$\|T(x) - T(y)\| \leq L\|x - y\| \quad \forall\, x, y \in \mathcal{H}$$

with $L < 1$.

If T is contractive, $x^k \to x^\star$ strongly with rate $\|x^k - x^\star\| \leq L^k\|x^0 - x^\star\|$.

(Banach contraction principle)

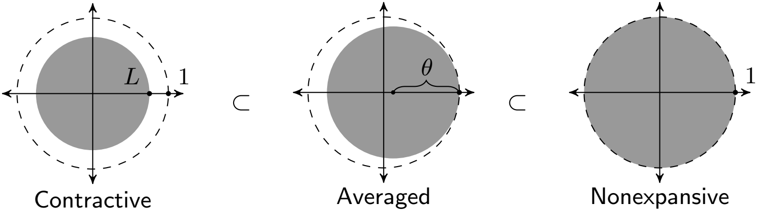
Convergence via operator properties: averaged

$T : \mathcal{H} \to \mathcal{H}$ is **averaged** if $T = (1-\theta)I + \theta R$ for some nonexpansive operator R and $\theta \in (0,1)$.

If T is averaged and has a fixed point, then $x^k \to x^\star$ weakly for some fixed point $x^\star$.

(Krasnosel'skiĭ–Mann theorem)

Convergence via operator properties



General rubric for proving convergence of a fixed-point iteration:

1. Prove T is contractive or averaged.
2. Apply convergence argument of Banach or Krasnosel'skiĭ–Mann.

Step 2 is routine. We present a **geometric** approach for step 1.

Outline

SRG of a nonlinear operator A

Pick $x \neq y$, $u \in Ax$, and $v \in Ay$. Consider the complex conjugate pair

$$z = \frac{\|u - v\|}{\|x - y\|} \exp\left[\pm i \angle(u - v, x - y)\right].$$

$\operatorname{Re} z$ and $\operatorname{Im} z$ respectively represent the components of $u - v$ aligned with and perpendicular to $x - y$, i.e.,

$$\operatorname{Re} z = \operatorname{sgn}(\langle u - v, x - y\rangle) \frac{\|P_{\operatorname{span}\{x-y\}}(u - v)\|}{\|x - y\|}$$

$$\operatorname{Im} z = \pm \frac{\|P_{\{x-y\}^\perp}(u - v)\|}{\|x - y\|}$$

Define the **scaled relative graph** (SRG) of $A : \mathcal{H} \rightrightarrows \mathcal{H}$ with

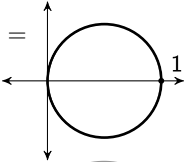
$$\mathcal{G}(A) = \{z \,|\, x \neq y, u \in Ax, v \in Ay\} \left(\cup \{\infty\} \text{ if } A \text{ is multi-valued} \right)$$

Interpretation: SRG to (nonlinear) operator $\cong$ eigenvalues to matrix.

Examples of $\mathcal{G}$

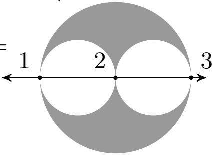
P_L = projection onto a line in $\mathbb{R}^2$:

$\mathcal{G}(P_L) =$



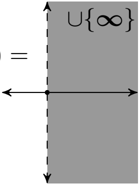
$$A(x_1, x_2, x_3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}:$$

$\mathcal{G}(A) =$



$\partial\| \cdot \|$ in $\mathbb{R}^n$, $n \geq 2$:

$\mathcal{G}(\partial\| \cdot \|) =$



Eigenvalues $\subseteq$ SRG

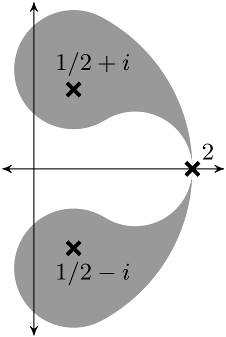
For matrices, the SRG generalizes eigenvalues.

Theorem

If $A \in \mathbb{R}^{n\times n}$ and $n = 1$ or $n \geq 3$,[2] *then $\Lambda(A) \subseteq \mathcal{G}(A)$.*

The figure shows SRG of

$$\begin{bmatrix} 1/2 & 2 & 0 \\ -1/2 & 1/2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

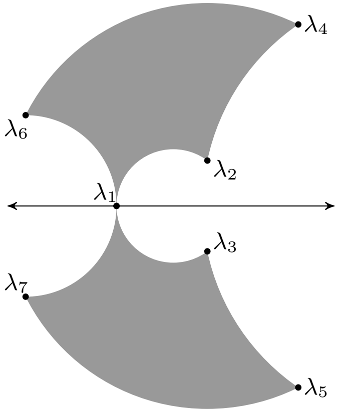


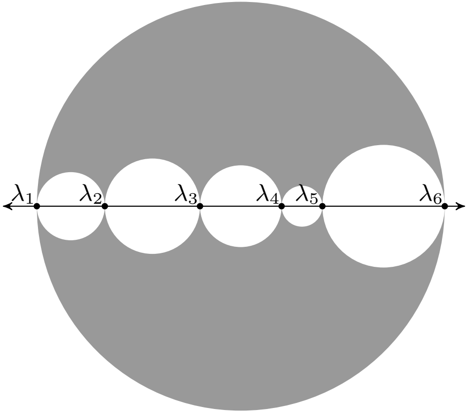
SRG is different from the numerical range (field of values) or the pseudospectrum.

[2] The result fails for $n = 2$ because S^{n-1}, the sphere in $\mathbb{R}^n$, is not simply connected for $n = 2$; the proof constructs a loop in S^{n-1} and argues the image of the loop on the complex plane is nullhomotopic.

SRG of normal matrices

SRG of normal matrices can be characterized with the Poincaré half-plane model of hyperbolic geometry [3].





[3] Huang, **Ryu**, Yin, Scaled Relative Graph of Normal Matrices, arXiv, 2019

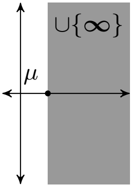
SRG of operator class $\mathcal{A}$

The SRG of an operator class $\mathcal{A}$ is defined by

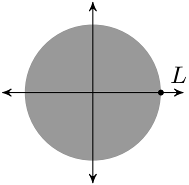
$$\mathcal{G}(\mathcal{A}) = \bigcup_{A \in \mathcal{A}} \mathcal{G}(A)$$

We focus on SRGs of operator classes, rather than SRGs of individual operators, because most theorems are stated with operator classes. E.g. "$I - A$ is nonexpansive if A is $1/2$-cocoercive."

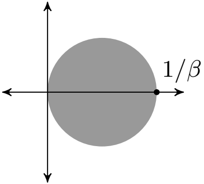
SRG of operator class $\mathcal{A}$



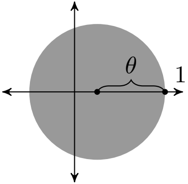
$\mathcal{M}_\mu$: μ-strongly monotone
$\partial\mathcal{F}_{\mu,\infty}$: gradient of μ-strgly-cvx. diff. func.



$\mathcal{L}_L$: L-Lipschitz



$\mathcal{C}_\beta$: β-cocoercive
$\partial\mathcal{F}_{0,\frac{1}{\beta}}$: gradient of $\frac{1}{\beta}$-Lip.diff.cvx.func.



$\mathcal{N}_\theta$: θ-averaged

Converse: from SRG to operator class

Given an operator class, we can draw the SRG, i.e.,

$$\text{operator class} \Rightarrow \text{SRG}$$

follows from the definition.

Conversely, can we look at an SRG and say something about the operator class? In general, no. To perform the reasoning

$$\text{SRG} \Rightarrow \text{operator class}$$

we need further conditions.

SRG-full classes

A class $\mathcal{A}$ of operators is **SRG-full** if

$$A \in \mathcal{A} \quad \Leftrightarrow \quad \mathcal{G}(A) \subseteq \mathcal{G}(\mathcal{A}).$$

i.e., membership of $\mathcal{A}$ is equivalent to containment of the SRG.
An SRG-full class is completely characterized by its SRG.

$A \in \mathcal{A} \Rightarrow \mathcal{G}(A) \subseteq \mathcal{G}(\mathcal{A})$ holds by definition of the SRG.
$A \in \mathcal{A} \Leftarrow \mathcal{G}(A) \subseteq \mathcal{G}(\mathcal{A})$ is the substance of this definition.

Theorem (Informal)

The important operator classes are SRG full.

Outline

Background and preliminaries

Scaled relative graph

Operator and SRG transformation

Operator tranformation $\cong$ SRG transformation

Algebraic operations on operators correspond to geometric operations on the SRG.

Under suitable conditions,

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

Use these to prove theorems.

Scaling and translation

Theorem

For $\alpha, \beta \in \mathbb{R}$ *and* $\alpha \neq 0$,

$$\mathcal{G}(\beta I + \alpha \mathcal{A}) = \beta + \alpha \mathcal{G}(\mathcal{A}).$$

Convergence analysis: gradient descent

Fact

Assume f is μ-strongly convex and L-smooth. Then

$$x^{k+1} = x^k - \alpha \nabla f(x^k)$$

converge exponentially to the minimizer $x^\star$ with rate

$$\|x^k - x^\star\| \leq \left(\max\{|1-\alpha\mu|, |1-\alpha L|\}\right)^k \|x^0 - x^\star\|.$$

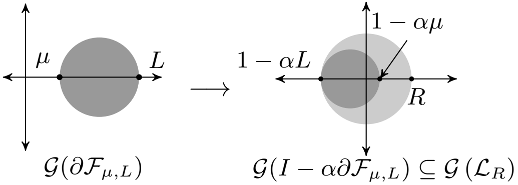
Proof. Theorem is equivalent to

$$I - \alpha \partial \mathcal{F}_{\mu,L} \subseteq \mathcal{L}_R$$

with $R = \max\{|1-\alpha\mu|, |1-\alpha L|\}$.

Inclusion of the class is equivalent to inclusion of the SRG

$$\mathcal{G}(I - \alpha\partial\mathcal{F}_{\mu,L}) \subseteq \mathcal{G}(\mathcal{L}_R).$$



□

Convergence analysis: forward step iteration

Fact

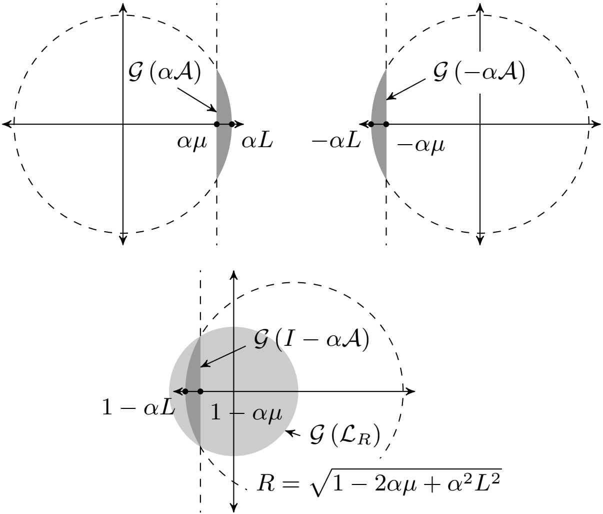
Assume A is μ-strongly monotone and L-Lipschitz. Then

$$x^{k+1} = x^k - \alpha A x^k$$

converge exponentially to the zero $x^\star$ with rate

$$\|x^k - x^\star\| \leq \left(1 - 2\alpha\mu + \alpha^2 L^2\right)^{k/2} \|x^0 - x^\star\|.$$

Proof.



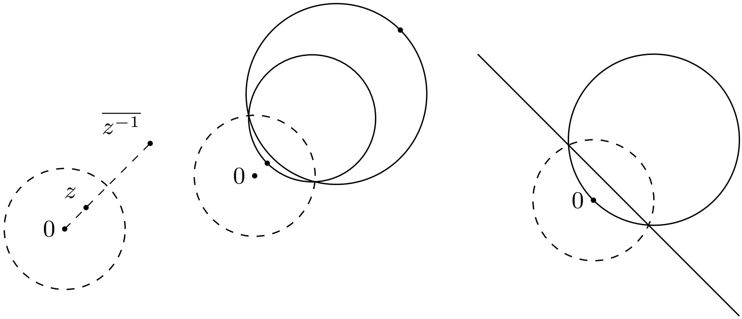
$\mathcal{G}(\alpha\mathcal{A})$
$\alpha\mu$
αL
$\mathcal{G}(-\alpha\mathcal{A})$
$-\alpha L$
$-\alpha\mu$
$\mathcal{G}(I-\alpha\mathcal{A})$
$1-\alpha L$
$1-\alpha\mu$
$\mathcal{G}(\mathcal{L}_R)$
$R=\sqrt{1-2\alpha\mu+\alpha^2L^2}$

□

Inversive geometry

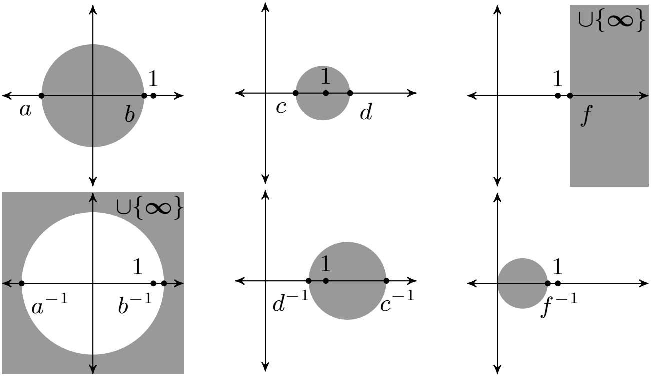
The **inversion** map is $z \mapsto \bar{z}^{-1}$. In polar form, $re^{i\varphi} \mapsto (1/r)e^{i\varphi}$, i.e., inversion preserves the angle and inverts the magnitude.

Inversion is a classical tool in Euclidean geometry, and is known as the Möbius transformation in complex analysis.



Inversive geometry

Generalized circles consist of a (finite) circles in $\overline{\mathbb{C}}$ and lines with $\{\infty\}$ in $\overline{\mathbb{C}}$. Inversion maps generalized circles to generalized circles.



Operator inversion $\cong$ SRG inversion

Theorem

$$\mathcal{G}(\mathcal{A}^{-1}) = \big(\mathcal{G}(\mathcal{A})\big)^{-1}.$$

(To clarify, $\big(\mathcal{G}(\mathcal{A})\big)^{-1} = \{z^{-1} \mid z \in \mathcal{G}(\mathcal{A})\}$.)

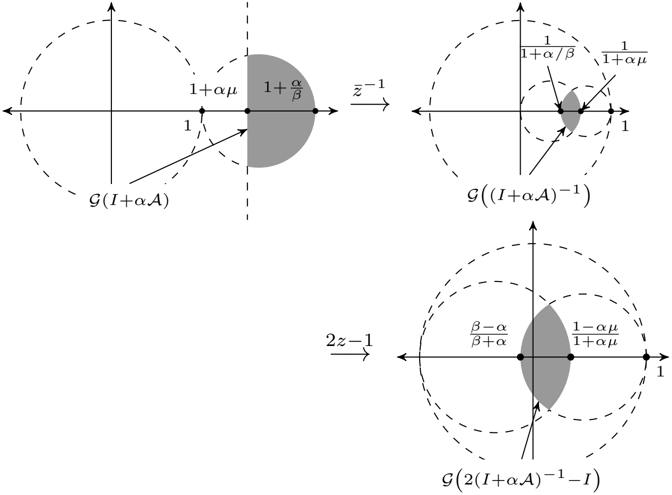
Convergence analysis: Peaceman–Rachford splitting

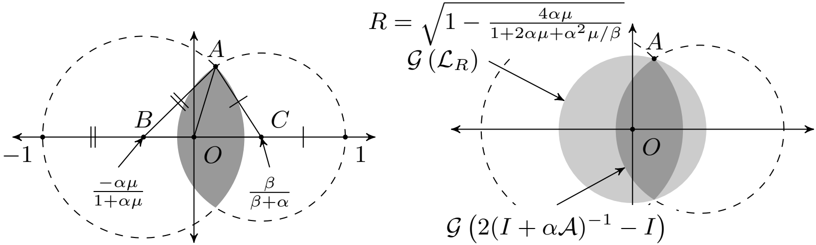
Fact

Assume A is μ-strongly monotone and β-cocoercive. Then $z^{k+1} = (2J_{\alpha A} - I)(2J_{\alpha B} - I)z^k$ converge exponentially to the fixed point $z^\star$ with rate

$$\|z^k - z^\star\| \leq \left(1 - \frac{4\alpha\mu}{1 + 2\alpha\mu + \alpha^2\mu/\beta}\right)^{k/2} \|z^0 - z^\star\|.$$

Proof.





By Stewart's theorem,

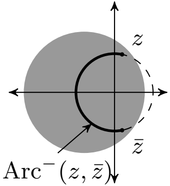
$$
\begin{aligned}
\overline{OA}^2 &= \frac{\overline{OC}\cdot\overline{AB}^2 + \overline{BO}\cdot\overline{CA}^2 - \overline{BO}\cdot\overline{OC}\cdot\overline{BC}}{\overline{BC}} \\
&= \frac{\frac{\beta}{\alpha+\beta}\left(1-\frac{\alpha\mu}{1+\alpha\mu}\right)^2 + \frac{\alpha\mu}{1+\alpha\mu}\left(1-\frac{\beta}{\alpha+\beta}\right)^2 - \frac{\beta}{\alpha+\beta}\frac{\alpha\mu}{1+\alpha\mu}\left(\frac{\beta}{\alpha+\beta}+\frac{\alpha\mu}{1+\alpha\mu}\right)}{\frac{\beta}{\alpha+\beta}+\frac{\alpha\mu}{1+\alpha\mu}} \\
&= 1 - \frac{4\alpha\mu}{1+2\alpha\mu+\alpha^2\mu/\beta}.
\end{aligned}
$$

□

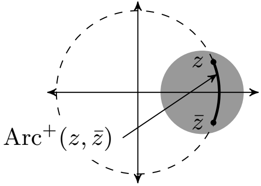
Composition of operators

Theorem

Let $\mathcal{A}$ and $\mathcal{B}$ be SRG-full classes. Assume the SRGs do not contain ∞ and are not empty. If $\mathcal{A}$ or $\mathcal{B}$ satisfies the left or right-arc property



Left arc property



Right arc property

then

$$\mathcal{G}(\mathcal{A}\mathcal{B}) = \mathcal{G}(\mathcal{B}\mathcal{A}) = \mathcal{G}(\mathcal{A})\mathcal{G}(\mathcal{B}).$$

(The SRGs commute even though the operators do not.)

Convergence: alternating projections

Fact

The alternating projections method $x^{k+1} = P_C P_D x^k$ *converge in that* $x^k \to x^\star$ *weakly for some* $x^\star \in C \cap D$.

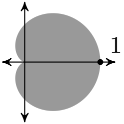
Follows from the following result.

Composition of firmly nonexpansive operators

Theorem

Let $\mathcal{N}_{1/2}$ be the class of firmly nonexpansive operators. Then

$$\mathcal{G}(\mathcal{N}_{1/2}\mathcal{N}_{1/2}) =$$

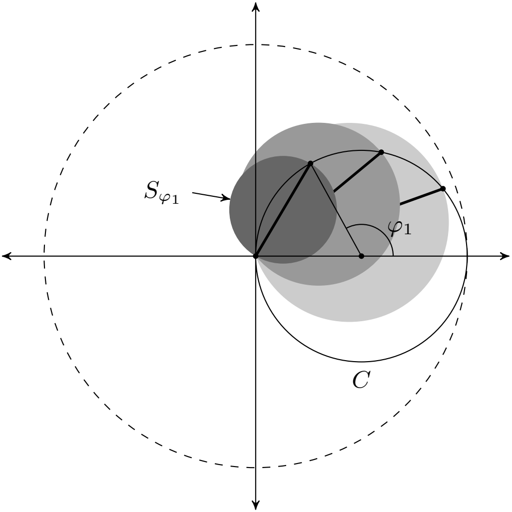


$$\left\{re^{i\varphi} \mid 0 \le r \le \cos^2(\varphi/2)\right\}$$

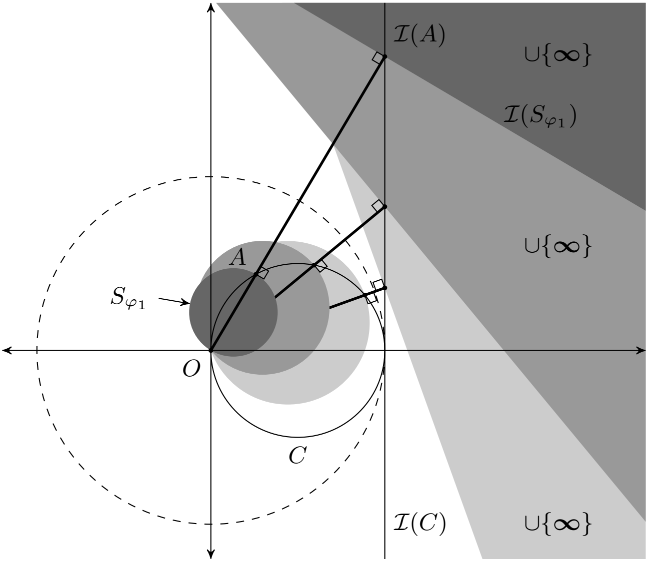
and $\mathcal{N}_{1/2}\mathcal{N}_{1/2} \subset \mathcal{N}_{2/3}$.

(Shape known as cardiod.)

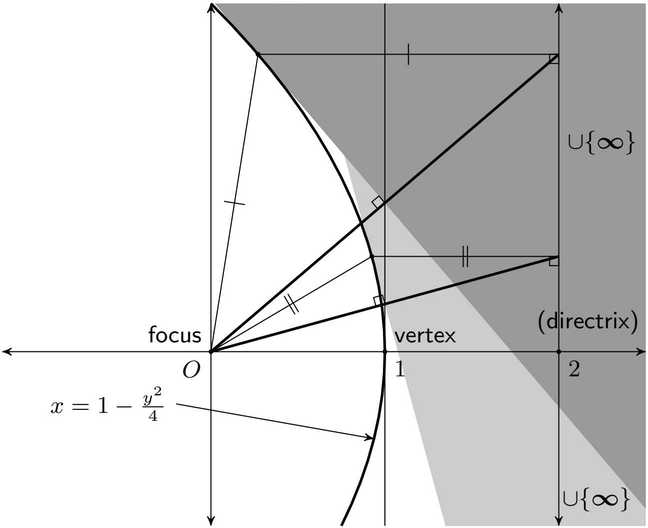
Proof outline. The SRG is the union with $\varphi_1 \in [0, 2\pi]$.



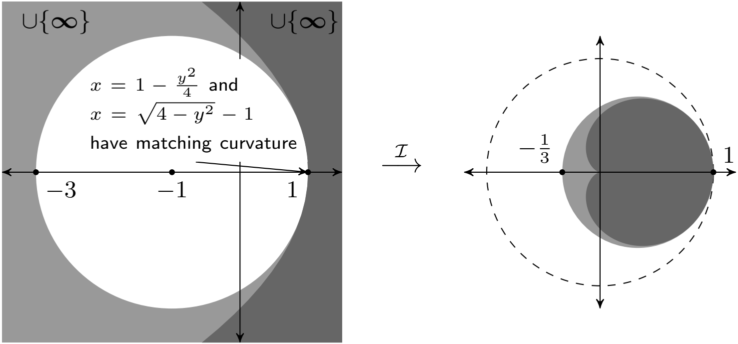
Write $\mathcal{I}$ for the inversion mapping. In the inverted space we have



The union of the half-spaces forms a parabola



Find the largest circle inscribed in the left of the parabola and invert.



□

Conclusion

- SRG maps the action of an operator to the 2D plane.
- Algebraic operations on operators correspond to geometric operations on SRGs.
- With SRG, we analyze fixed-point iterations with geometric proofs.
- SRG has also been used to establish convergence of the deep-learning based "Plug-and-Play" method for image denoising (ICML 2019).

References:

- **Ryu**, Hannah, Yin, Scaled Relative Graph: Nonexpansive operators via 2D Euclidean Geometry, arXiv, 2019.
- Huang, **Ryu**, Yin, Tight Coefficients of Averaged Operators via Scaled Relative Graph, Journal of Mathematical Analysis and Applications, 2020.
- Huang, **Ryu**, Yin, Scaled Relative Graph of Normal Matrices, arXiv, 2019.
- **Ryu**, Liu, Wang, Chen, Wang, Yin, Plug-and-Play Methods Provably Converge with Properly Trained Denoisers, ICML, 2019.