NORTH-HOLLAND

Shorted Operators: An Application in Potential Theory

Volker Metz*
Department of Mathematics
University of California San Diego
La Jolla, California 92093-0112

Submitted by Simo Puntanen

ABSTRACT

On nested fractals a "Laplacian" can be constructed as a scaled limit of difference operators. The appropriate scaling and starting configuration are given by a nonlinear, finite dimensional eigenvalue problem. We study it as a fixed point problem using Hilbert's projective metric on cones, a nonlinear generalization of the Perron-Frobenius theory of nonnegative matrices. The nonlinearity arises from a map Φ known as the shorted operator. Potential theoretic notions and results apply to it, since it acts on a cone of discrete "Laplacians" or difference operators. Usually, Φ is considered on the larger cone of positive semidefinite operators. We are able to take advantage of the more specific structure of the reduced domain because several properties of Φ are local. Results are possible with respect to continuity, concavity, the Fréchet derivative, invariant subcones, the geometry of these cones, and the contraction of Hilbert's metric. © 1997 Elsevier Science Inc.

---

1. INTRODUCTION

We will deal with the matrix analytic aspects of a nonlinear eigenvalue problem whose physical and mathematical background can be found in [4]. Mathematically it arises in the construction of a "Laplacian" on nested Fractals. Physically it is a renormalization problem. The troublesome ingredi-

---

*Current address: Faculty of Mathematics, University of Bielefeld, 33501 Bielefeld, Germany. E-mail: `metz@mathematik.uni-bielefeld.de`.

LINEAR ALGEBRA AND ITS APPLICATIONS 264:439–455 (1997)

655 Avenue of the Americas, New York, NY 10010
0024-3795/97/$17.00
PII S0024-3795(96)00303-5

ent is the shorted operator, which is a version of the Schur complement. We will see below that in potential theory it is known as "traces of Dirichlet forms" or "traces of Markov processes."

To simplify the notation and clarify the application we consider a specific nested fractal, the Vicsek snowflake X. It can be constructed by the five similitudes $\psi_1, \dots, \psi_5 : \mathbb{R}^2 \to \mathbb{R}^2$ given below:

$$\psi_1(x) := \frac{x}{3} + \frac{2}{3}(1,1), \qquad \psi_2(x) := \frac{x}{3} + \frac{2}{3}(0,1), \qquad \psi_3(x) := \frac{x}{3},$$

$$\psi_4(x) := \frac{x}{3} + \frac{2}{3}(1,0), \qquad \psi_5(x) := \frac{x}{3} + \frac{1}{3}(1,1).$$

They contract the unit square $[0,1]^2$ by $\frac{1}{3}$ and arrange five such small copies in a chessboard pattern inside $[0,1]^2$. We define $\Psi(M) := \bigcup_{i=1}^{5}\psi_i(M)$ for $M \subset \mathbb{R}^2$. Now the fractal is defined by $X := \bigcap_{i \in \mathbb{N}}\Psi^i([0,1]^2)$. Each ψ_i has a unique fixed point x_i, namely,

$$x_1 = (1,1), \quad x_2 = (0,1), \quad x_3 = (0,0), \quad x_4 = (1,0), \quad x_5 = \left(\tfrac{1}{2}, \tfrac{1}{2}\right).$$

Set $F_0 := \{x_1, \dots, x_4\}$ and $F_i := \Psi^i(F_0)$ for all $i \in \mathbb{N}$. The set $\bigcup_{i \in \mathbb{N}} F_i$ is dense in X with respect to the Euclidean topology on $\mathbb{R}^2$. On X we consider the $(\ln 5)/(\ln 3)$-dimensional Hausdorff measure μ with $\mu(X) = 1$ (cf. [17]). The fractal X is invariant under every reflection $\rho_{x,y}$ in the hyperplane of points equidistant from x and y, for all points $x, y \in F_0$ with $x \neq y$. Let $\mathfrak{G}$ denote the group generated by these reflections.

The eigenvalue problem will be formulated in terms of difference forms or (potential theoretically) of *Dirichlet forms*, the analog of the classical Dirichlet integral [10]. For $i \in \{0,1\}$ let us define $\mathcal{D}_i := \{f \mid f : F_i \to \mathbb{R}\}$ and denote the Euclidean inner product on $\mathcal{D}_i$ by $\langle \cdot, \cdot \rangle_i$. Let 1_M be the characteristic function of the set $M \subset F_1$. We will regard 1_M as an element of $\mathcal{D}_0$ and $\mathcal{D}_1$. We define a Dirichlet form $\mathcal{E}$ with domain $\mathcal{D}_i$ on the Hilbert space $(\mathcal{D}_i, \langle \cdot, \cdot \rangle_i)$ with $\mathcal{E}(1_{F_i}, 1_{F_i}) = 0$ as follows:

$$\mathcal{E}(f,g) = \frac{1}{2} \sum_{x,y \in F_i} [f(y) - f(x)][g(y) - g(x)] c_{\mathcal{E}}(x,y), \qquad (1.1)$$

for all $f, g \in \mathcal{D}_i$ and a unique function (conductance) $c_{\mathcal{E}} : F_i^2 \to \mathbb{R}_+$ which is symmetric and vanishes on the diagonal. We call the operator defined by $\mathcal{E}$ a *Dirichlet operator*. It is the analog of the classical Laplace operator, and in our case it is also a difference operator. On the other hand, a Dirichlet form that vanishes on constants defines a unique conductance (cf. [20, Section 2]). The isomorphism between conductances and Dirichlet forms will be denoted by Ξ. We will only consider Dirichlet forms $\mathscr{E}$ and conductances c which are invariant under $\mathfrak{G}$, that is, $\mathscr{E}(f \circ \rho, g \circ \rho) = \mathscr{E}(f, g)$ and $c(\rho(x), \rho(y)) = c(x, y)$ for all $f, g \in \mathscr{D}_i$, $\rho \in \mathfrak{G}$, and $x, y \in F_0$. The cone of all such Dirichlet forms on $\mathscr{D}_0$ is denoted by $\mathbb{D}$. A cone $\mathbb{P}$ of positive semidefinite forms on $\mathscr{D}_0$ is defined by

$$\mathbb{P} := \{\mathscr{E} - \mathscr{F} | \mathscr{E}, \mathscr{F} \in \mathbb{D},\ \mathscr{E}(f,f) \geqslant \mathscr{F}(f,f) \text{ for all } f \in \mathscr{D}_0\}.$$

An embedding Hilbert space $(\mathbb{B}, \langle \cdot , \cdot \rangle)$ is given by $\mathbb{B} := \mathbb{D} - \mathbb{D}$ and $\langle \mathscr{E}, \mathscr{F} \rangle :=$ trace$(E \circ F)$, where E, F are the Dirichlet operators of $\mathscr{E}$ and $\mathscr{F}$ respectively. For $\mathscr{F} - \mathscr{E} \in \mathbb{P}$ we write $\mathscr{E} \leqslant \mathscr{F}$. Since $\mathscr{E}$ and $\mathscr{F}$ are symmetric, this coincides with $\mathscr{E}(f,f) \leqslant \mathscr{F}(f,f)$ for all $f \in \mathscr{D}_0$, the Loewner ordering. The norm on $\mathbb{B}$ is monotone, that is, $\langle \mathscr{E}, \mathscr{E} \rangle \leqslant \langle \mathscr{F}, \mathscr{F} \rangle$ if $\mathscr{E} \leqslant \mathscr{F}$.

In our terminology the eigenvalue problem can be formulated as follows. For $\mathscr{E} \in \mathbb{B}$ and $f, g \in \mathscr{D}_1$ we define the linear coupling map Ψ by

$$\Psi(\mathscr{E})(f, g) := \sum_{i=1}^{5} \mathscr{E}(f \circ \psi_i, g \circ \psi_i).$$

This defines a Dirichlet form $\Psi(\mathscr{E})$ on $\mathscr{D}_1$ with $\Psi(\mathscr{E})(1_{F_1}, 1_{F_1}) = 0$. Graphically Ψ assembles a refined "grid" F_1 from five coupled copies of the initial "grid" F_0. By the definition of our fractal, $\Psi(\mathscr{E})$ is also invariant under $\mathfrak{G}$. In the next step we eliminate the vertices $F_1 \setminus F_0$ from the fine "grid" F_1 to relate it to a coarse "grid" F_0. For $f \in \mathscr{D}_0$ and $\mathscr{E} \in \mathbb{P}$ we define the nonlinear "trace" map Φ [10, Section 6.2] by

$$\Phi(\Psi(\mathscr{E}))(f,f) := \inf\{\Psi(\mathscr{E})(g, g) | g \in \mathscr{D}_1,\ g|_{F_0} = f\}.$$

In physics the action of the map Φ is called "coarse graining renormalization" [11]. In our terms it is the shorted operator. In the setup of [2, Theorem 1] the operator to be shorted is $\Psi(E)$, the operator of our Dirichlet form $\Psi(\mathscr{E})$; the underlying cone of operators is the set of all operators which define a form in $\Psi(\mathbb{P})$ equipped with the Loewner ordering $\leqslant$; and the column range of the shorted operator $\Phi(\Psi(E))$ has to equal $\mathscr{D}_0$ modulo the constants.

Now, back to our terminology. Let $f \in \mathscr{D}_0$ and $\mathscr{E} \in \mathbb{P}^0$, the interior of $\mathbb{P}$ in $(\mathbb{B}, \langle \cdot , \cdot \rangle^{1/2})$. Then the above variational problem has a unique minimizing element $H_{\mathscr{E}} f \in \mathscr{D}_1$. When $\mathscr{E}$ is an element of $\partial \mathbb{P}$, the boundary of $\mathbb{P}$, there are several minimizing elements. Let $H_{\mathscr{E}} f$ denote the unique one with minimal $\langle \cdot , \cdot \rangle_2^{1/2}$-norm. The linear map $f \mapsto H_{\mathscr{E}} f$ will be denoted by $H_{\mathscr{E}} : \mathscr{D}_0 \to \mathscr{D}_1$. Thus,

$$\Phi(\Psi(\mathscr{E}))(f, f) = \Psi(\mathscr{E})(H_{\mathscr{E}} f, H_{\mathscr{E}} f).$$

For $\mathscr{E} \in \mathbb{D}$, the map $H_{\mathscr{E}}$ is known in potential theory as the "harmonic kernel" of $\Psi(\mathscr{E})$ on $F_1 \setminus F_0$. The minimal Euclidean norm in its definition guarantees a consistent probabilistic interpretation [6, Section 1.6]. The above equation is known as the *Dirichlet principle*. By this principle or by [2, Theorem 1] we have $H_{\mathscr{E}} f|_{F_0} = f$ and $E H_{\mathscr{E}} f|_{F_1 \setminus F_0} \equiv 0$. This means that $H_{\mathscr{E}} f$ solves a "Dirichlet problem" on $F_1 \setminus F_0$ with boundary data f on F_0.

An explicit version of the above formula is given in [1, Formula 7]: Let E' denote the operator of $\Psi(\mathscr{E})$, and $\Phi(E')$ the operator of $\Phi(\Psi(\mathscr{E}))$. Furthermore, let π be the orthogonal projection onto $\mathscr{D}_0$ with respect to the Euclidean inner product on $\mathscr{D}_1$, and π_o the corresponding projection onto the orthogonal complement of $\mathscr{D}_0$. The adjoint is denoted by $(\cdot)^*$. Then

$$\Phi(E') = \pi E' \pi^* - \pi E' \pi_o^* (\pi_o E' \pi_o^*)^+ \pi_o E' \pi^*, \tag{1.2}$$

where A^+ is the Moore-Penrose generalized inverse of the matrix A [24]. The right hand side of this equation is a "generalized Schur complement" [5, Formula 11].

Finally we set $\Lambda := \Phi \circ \Psi$, and our eigenvalue problem is

$$\exists \gamma > 0 \ \exists \mathscr{E} \in \mathbb{D} \cap \mathbb{P}^\circ : \Lambda(\mathscr{E}) = \gamma \mathscr{E}. \tag{1.3}$$

Note that $\ker \mathscr{E}$ equals the constants if and only if $\mathscr{E} \in \mathbb{P}^\circ$. So $\mathscr{E} \in \mathbb{D} \cap \mathbb{P}^\circ$ means that $\mathscr{E}$ is an *irreducible* Dirichlet form, that is, the underlying graph $(F_0, \{\{x, y\} \mid c(x, y) > 0\})$ is connected. The existence of a solution of (1.3) was shown in [17, Theorem V.5] and its uniqueness was recently proved in [26, Théorème V.2].

Our eigenvalue problem (1.3) fits nicely into a fixed point theory known as *Hilbert's projective metric on cones* (cf. [23]). For $\mathscr{A}, \mathscr{B} \in \mathbb{P}^\circ$ we define Hilbert's projective metric h by

$$\lceil \mathscr{A} / \mathscr{B} \rceil := \inf\{\alpha > 0 \mid \mathscr{A} \leqslant \alpha \mathscr{B}\},$$

$$\lfloor \mathscr{A} / \mathscr{B} \rfloor := \sup\{\alpha > 0 \mid \alpha \mathscr{B} \leqslant \mathscr{A}\},$$

$$h(\mathscr{A}, \mathscr{B}) := \ln \frac{\lceil \mathscr{A} / \mathscr{B} \rceil}{\lfloor \mathscr{A} / \mathscr{B} \rfloor}.$$

Dividing $\mathscr{D}_0$ by the constants, a simultaneous diagonalization of $\mathscr{A}$ and $\mathscr{B}$ shows that $[\mathscr{A}/\mathscr{B}]$ is the largest eigenvalue of $B^{-1}A$, where A and B are the operators of the forms $\mathscr{A}$ and $\mathscr{B}$ respectively (cf. Proposition 2.7). The function h differs from a metric in the following respects: $h(\mathscr{A}, \mathscr{B}) = 0$ if and only if $\mathscr{A}$ and $\mathscr{B}$ lie on the same ray; for all strictly positive α and β we have $h(\alpha\mathscr{A}, \beta\mathscr{B}) = h(\mathscr{A}, \mathscr{B})$. Let $S_1(0) := \{\mathscr{B} \in \mathbb{B} \mid \|\mathscr{B}\| = 1\}$. Then $(\dot{S}_1(0) \cap \mathbb{P}^\circ, h)$ is a complete metric space [23, Theorem 1.2]. An application of Hilbert's metric to our example in [21, Formula 4.2] results in: For all $\mathscr{A}, \mathscr{B} \in \mathbb{P}^\circ$ there exists a $q(\mathscr{A}, \mathscr{B}) \in (0, 1)$ such that for all $n \in \mathbb{N}$,

$$h(\Lambda^n(\mathscr{A}), \Lambda^n(\mathscr{B})) \leqslant 2 \cdot q(\mathscr{A}, \mathscr{B})^n \cdot h(\mathscr{A}, \mathscr{B}). \tag{1.4}$$

If $\mathscr{A}$ tends to $\partial\mathbb{D} \cap \partial\mathbb{P}$ then $q(\mathscr{A}, \mathscr{B})$ tends to 1. This convergence result not only implies existence and uniqueness for the original eigenvalue problem (1.3), but it also produces an approximation to the solution by scaled iteration of Λ. Numerically this is done with the help of (1.2).

Section 2 contains various results obtained by studying the action of Λ on $\mathbb{D}$. The connections between central potential theoretic concepts and the shorted operator are explained in Section 3. The effective resistance from electrical network theory is also considered in this context. The interpretation of (1.3) as a dynamical system is the main difference between our use of the shorted operator and the setup in [2]. The iteration of Λ/γ can be interpreted as the shorting of an infinite model or potential theoretically as homogenization of the fractal's "Laplacian." This is discussed in Section 4.

2. SELECTED BENEFITS FROM $\mathbb{D} \subset \mathbb{P}$

The following selected list of results on Λ rely on the fact that $\mathbb{D}$ has a more specific structure than $\mathbb{P}$. Let us denote the infimum of a function f and the constant function 1 by $f \wedge 1$. With this notation we can express the so-called Markov property of Dirichlet forms:

$$\mathbb{D} = \{\mathscr{E} \in \mathbb{P} \,|\, \mathscr{E}(f \wedge 1, f \wedge 1) \leqslant \mathscr{E}(f, f) \text{ for all } f \in \mathscr{D}_0\}.$$

In terms of linear algebra, $\mathbb{D}$ contains those forms of $\mathbb{P}$ that are also M-matrices. It is the specific sign structure of these M-matrices that we are going to take advantage of (cf. [13, Section 2.5]). In probabilistic potential theory the elements of $\mathbb{D}$ correspond to "symmetric Markov processes" with the (one step) transition probability

$$P(x,y) := \frac{c(x,y)}{c(x)} \quad \text{with} \quad c(x) := \sum_{z \in F_0} c(x,z)$$

for all $x, y \in F_0$ [6, Section 2.6]. At the same time an element of $\mathbb{D}$ corresponds to an electrical resistor network given by the terminals F_0 and the conductances $c(x,y)$ on the edges $\{\{x,y\} \subset F_0^2 \mid c(x,y) > 0\}$ as described in [6, Section 3.1]. In the present example the analytic terminology turned out to be very effective, whereas the physical and probabilistic view gave much of the intuition. The latter approach is especially well suited for pathwise arguments on a network.

In our specific example, $\mathbb{B} \simeq \mathbb{R}^2$, since $\mathfrak{G}$ allows only two different conductances. The set $\mathbb{D}$ can be identified with $\mathbb{R}_+^2$ by defining the first component of (a,b) to denote the conductance on the sides of F_0 and the second component to give the conductance on the diagonals of F_0. Furthermore, $\mathbb{P} \simeq \{(a,b) \in \mathbb{R}^2 \mid a \geqslant 0,\ a+b \geqslant 0\}$, since the Dirichlet form defined by (a,b) has an eigenvalue 0, another eigenvalue $4a$, and twice the eigenvalue $2(a+b)$ (cf. Proposition 2.7).

PROPOSITION 2.1 [19, Theorem 2.2].

(a) $\Lambda(\mathbb{D}) \subset \mathbb{D}$.
(b) $\Lambda \in \mathscr{C}(\mathbb{D})$.

Proof. (a): Because of the Dirichlet principle and the Markov property we derive

$$\begin{aligned}\Lambda(\mathscr{A})(f \wedge 1, f \wedge 1) &= \Psi(\mathscr{A})(H_{\mathscr{A}}(f \wedge 1), H_{\mathscr{A}}(f \wedge 1))\\ &\leqslant \Psi(\mathscr{A})((H_{\mathscr{A}}f) \wedge 1, (H_{\mathscr{A}}f) \wedge 1)\\ &\leqslant \Psi(\mathscr{A})(H_{\mathscr{A}}f, H_{\mathscr{A}}f)\\ &= \Lambda(\mathscr{A})(f,f).\end{aligned}$$

This proves the Markov property of $\Lambda(\mathscr{A})$.

(b): Consider $\mathscr{A}, \mathscr{B} \in \mathbb{D}$ and $f \in \mathscr{D}_0$. Again by the Dirichlet principle

$$\begin{aligned}\Lambda(\mathscr{A})(f,f) - \Lambda(\mathscr{B})(f,f) &= \Psi(\mathscr{A})(H_{\mathscr{A}}f, H_{\mathscr{A}}f) - \Psi(\mathscr{B})(H_{\mathscr{B}}f, H_{\mathscr{B}}f)\\ &\leqslant \Psi(\mathscr{A})(H_{\mathscr{B}}f, H_{\mathscr{B}}f) - \Psi(\mathscr{B})(H_{\mathscr{B}}f, H_{\mathscr{B}}f)\\ &\leqslant |||\Psi||| \cdot \|\mathscr{A} - \mathscr{B}\| \cdot \|H_{\mathscr{B}}f\|_2^2,\end{aligned}$$

where $|||\cdot|||$ denotes the operator norm. The minimum principle for the harmonic function $H_{\mathscr{B}}f$ (cf. [20, Proposition 2.4]) gives us

$$\begin{aligned}\|H_{\mathscr{B}}f\|_2^2 &= \sum_{x \in F_0} (H_{\mathscr{B}}f)^2(x)\\ &\leqslant \|f\|_\infty^2 \cdot |F_1|,\end{aligned}$$

where $|F_1|$ denotes the number of elements in F_1, and $\|\cdot\|_\infty$ the sup norm. A similar argument for $\Lambda(\mathscr{B})(f,f) - \Lambda(\mathscr{A})(f,f)$ results in the same bound. Putting these facts together, we arrive at

$$|\Lambda(\mathscr{A})(f,f) - \Lambda(\mathscr{B})(f,f)| \leqslant |||\Psi||| \cdot \|f\|_\infty^2 \cdot |F_1| \cdot \|\mathscr{A} - \mathscr{B}\|.$$

This proves the desired continuity. ■

The continuity properties of Λ on $\mathbb{P}$ are weaker in general [2, p. 63]. Our next result complements Proposition 2.1(a).

PROPOSITION 2.2. $\Lambda(\mathbb{D} \cap \mathbb{P}^\circ) \subset \mathbb{D}^\circ$.

Proof. As an abbreviation of $1_{\{x\}}$ let us use δ_x. Because $\Xi(\mathbb{R}_+^2) = \mathbb{D}$, we know that $\mathscr{E} \in \mathbb{D}^\circ$ if and only if $\mathscr{E}(\delta_x, \delta_y) < 0$ for all $x, y \in F_0$ with $x \neq y$. Therefore, let $\mathscr{A} \in \mathbb{D} \cap \mathbb{P}^\circ$ and $x, y \in F_0$ with $x \neq y$. We denote the operator of $\Psi(\mathscr{A})$ by $\Psi(A)$. By the definition of Λ and the fact that $\Psi(A)H_{\mathscr{A}}\delta_x|_{F_1 \setminus F_0} \equiv 0$ we compute

$$\begin{aligned}\Lambda(\mathscr{A})(\delta_x, \delta_y) &= \Psi(\mathscr{A})(H_{\mathscr{A}}\delta_x, H_{\mathscr{A}}\delta_y)\\ &= \Psi(\mathscr{A})(H_{\mathscr{A}}\delta_x, \delta_y)\\ &= \Psi(A)H_{\mathscr{A}}\delta_x(y).\end{aligned}$$

Since $\mathscr{A} \in \mathbb{D} \cap \mathbb{P}^\circ$, it is irreducible. The minimum principle and $x \neq y$ now imply $\Psi(A)H_{\mathscr{A}}\delta_x(y) < 0$. ■

Lindstrøm used probabilistic pathwise arguments to prove for all nested fractals that the cone $\mathbb{L} \subset \mathbb{P}^\circ$ with

$$\mathbb{L} := \{(a, b) \in \mathbb{R}_+^2 \,|\, a \geqslant b\}$$

is also Λ-invariant. Together with the Propositions 2.1, 2.2 and Brouwer's fixed point theorem, this implies the following result.

PROPOSITION 2.3 [17, Theorem V.5]. *There exists an* $\mathscr{F} \in \mathbb{D}^\circ$ *and a* $\gamma > 0$ *such that* $\Lambda(\mathscr{F}) = \gamma\mathscr{F}$.

It is well known that Λ is positively homogeneous and $\mathbb{P}$-monotone [3, Theorems 10, 5]. This implies, as in [11, Corollary 3.7], that if $\mathscr{E} \in \mathbb{P}^\circ$ and $\alpha > 0$ with $\Lambda(\mathscr{E}) = \alpha\mathscr{E}$, then $\alpha = \gamma$. Therefore, we define

$$\mathrm{Fix} := \{\mathscr{E} \in \mathbb{P}^\circ | \Lambda(\mathscr{E}) = \gamma\mathscr{E}\}.$$

COROLLARY 2.4. $\mathrm{Fix} \subset \mathbb{D}^\circ$.

Proof. Dividing the eigenvalue equation by γ, we get a fixed point equation. The corresponding fixed point problem on $\mathbb{P}^\circ$ equipped with a close relative of Hilbert's metric allows us to conclude that Fix is arcwise connected, since $\mathbb{P}^\circ$ is [19, Corollary 4.12]. Now assume that there exists an $\mathscr{A} \in \mathrm{Fix} \cap (\mathbb{P}^\circ \setminus \mathbb{D}^\circ)$. Then the arc of fixed points in $\mathbb{P}^\circ$ that connects $\mathscr{A}$ and $\mathscr{F}$, which exists by Proposition 2.3, must intersect $\partial\mathbb{D} \cap \mathbb{P}^\circ$. But according to Proposition 2.2 there can be no fixed point in this set. ■

Recently, Sabot proved the uniqueness of eigenvectors of Λ for all nested fractals.

PROPOSITION 2.5 [26, Théorème V.2]. *Let* $\mathscr{F}$ *be as in Proposition* 2.3. *Then* $\mathrm{Fix} = \{\alpha\mathscr{F} | \alpha > 0\}$.

This result relies heavily on a specific decomposition of an irreducible Dirichlet form $\mathscr{A}$ into its trace $\mathscr{A}_t$ on a given subset M and its complement $\mathscr{A}_p$ [2, Theorem 2]. The form $\mathscr{A}_p$ is the reason why the Schur complement is called a complement and it is even better known in potential theory than the trace. It is referred to as the "part" of the Dirichlet form $\mathscr{A}$ on the complement of M. In probability theory defining parts is known as "killing of Markov processes" [10, Theorem A.2.10].

A more detailed analysis of our example shows that the source of its nice h-contraction in (1.4) is its strict superadditivity in certain directions. In this sense the concavity of the shorting operation is crucial, and (1.4) is partly a concavity statement. For $\mathscr{A}, \mathscr{B} \in \mathbb{P}^{\circ}$ let us define

$$\mathscr{C} := \mathscr{A} - \lfloor \mathscr{A}/\mathscr{B} \rfloor \mathscr{B} \in \partial \mathbb{P}.$$

According to [2, Theorem 4] the trace map is superadditive, hence Λ is, and there exists an $\mathscr{R} \in \mathbb{P}$ such that

$$\Lambda(\mathscr{A}) = \Lambda(\lfloor \mathscr{A}/\mathscr{B} \rfloor \mathscr{B}) + \Lambda(\mathscr{C}) + \mathscr{R}.$$

Since $\Lambda$is positively homogeneous, we conclude

$$\Lambda(\mathscr{A}) = \lfloor \mathscr{A}/\mathscr{B} \rfloor \Lambda(\mathscr{B}) + \Lambda(\mathscr{C}) + \mathscr{R}.$$

A similar equation can be derived for $\lceil \mathscr{A}/\mathscr{B} \rceil$. Together with the definition of h we see that Λ contracts h if it is strictly positive [that is, $\Lambda(\mathbb{P}) \subset \mathbb{P}^{\circ}$] and/or it is strictly superadditive [that is, $\Lambda(\mathscr{C} + \mathscr{B})(f, f) > \Lambda(\mathscr{C})(f, f) + \Lambda(\mathscr{B})(f, f)$ for all $f \in \mathscr{D}_0$ and all linearly independent $\mathscr{C}, \mathscr{B}$]. Concavity can be stated in terms of the first derivative. So let us have a look at the Fréchet derivative of Λ.

The cone $\mathbb{P}$ is embedded in a Banach space $\mathbb{B}$ and it is known that $D\Lambda(\mathscr{A})(\mathscr{B})$, the Fréchet derivative of Λ at $\mathscr{A} \in \mathbb{P}^{\circ}$ applied to $\mathscr{B} \in \mathbb{B}$, is given by $\Psi(\mathscr{B})(H_{\mathscr{A}} \cdot, H_{\mathscr{A}} \cdot)$ because of the corresponding formula for the derivative of the trace map Φ [3, Theorem 12]. A first consequence is $D\Lambda(\mathscr{A})(\mathbb{P}) \subset \mathbb{P}$.

Proposition 2.6. *Let $\mathscr{A} \in \mathbb{D}^{\circ}$. Then $D\Lambda(\mathscr{A})(\mathbb{D} \setminus \{0\}) \subset \mathbb{P}^{\circ}$.*

Proof. Let $\mathscr{A} \in \mathbb{D}^{\circ}$, $\mathscr{B} \in \mathbb{D} \setminus \{0\}$, and $f \in \mathscr{D}_0$ not constant. Then

$$\begin{aligned}[][D\Lambda(\mathscr{A})(\mathscr{B})](f, f) &= \Psi(\mathscr{B})(H_{\mathscr{A}} f, H_{\mathscr{A}} f) \\ &= \sum_{i=1}^{5} \mathscr{B}(H_{\mathscr{A}} f \circ \psi_i, H_{\mathscr{A}} f \circ \psi_i).\end{aligned}$$

Choose $x \in F_0$ such that $f(x) = \max_{y \in F_0} f(y)$. There exists $1 \leqslant j \leqslant 5$ with $\psi_j(x) = x$. Define $\Xi^{-1}(\mathscr{B}) =: c$. Then

$$\mathscr{B}(H_{\mathscr{A}}f \circ \psi_j, H_{\mathscr{A}}f \circ \psi_j) = \frac{1}{2} \sum_{y, z \in F_0} \left[H_{\mathscr{A}}f \circ \psi_j(y) - H_{\mathscr{A}}f \circ \psi_j(z)\right]^2 c(x, y).$$

Since $\mathscr{A}$ is irreducible, the minimum principle for the harmonic function $H_{\mathscr{A}}f$ and $\psi_j(F_0) \cap F_0 = \{x\}$ imply

$$f(x) = H_A f \circ \psi_j(x) > H_{\mathscr{A}}f \circ \psi_j(y) \qquad (y \in F_0 \setminus \{x\}).$$

By definition $\mathfrak{G}$ acts transitively on F_0. Thus $c(x) = 0$ implies $\mathscr{B} = 0$. Combining these facts, we arrive at $[D\Lambda(\mathscr{A})(\mathscr{B})](f, f) > 0$. Hence the kernel of $D\Lambda(\mathscr{A})(\mathscr{B})$ consists only of constants, that is, $D\Lambda(\mathscr{A})(\mathscr{B}) \in \mathbb{P}^{\circ}$. ∎

In general it is not true that $D\Lambda(\mathscr{A})(\mathbb{P}) \subset \mathbb{P}^{\circ}$, although $D\Lambda(\mathscr{A})(\mathbb{P}^{\circ}) \subset \mathbb{P}^{\circ}$. The behavior of $D\Lambda$ is for example important for the uniqueness of eigenvectors of Λ [19, Theorem 4.9] or the contraction of Hilbert's projective metric in certain regions [22, Proposition 3.3].

The cone $\mathbb{D}$ is always polyhedral, that is, the intersection of finitely many half spaces. Its extremal rays are spanned in our case by those forms corresponding to $(1, 0)$ and $(0, 1)$. It is not obvious that the cone $\mathbb{P}$ is also polyhedral. Proposition 2.7 makes use of the connection between Dirichlet forms and adjacency matrices of graphs and is valid for every nested fractal.

PROPOSITION 2.7 [21, Proposition 3.2]. *All elements of* $\mathbb{B}$ *commute, and* $\mathbb{P}$ *is a simplicial cone in* $\mathbb{B}$.

Proof. A conductance c on F_0 defines a graph $\Gamma(c)$ with vertex set F_0 and edge set $\{\{x, y\} \subset F_0 \mid c(x, y) > 0\}$. Hence, $\Gamma(1)$ is the complete graph with vertices F_0. To each graph $\Gamma = (F_0, E)$ we can associate a conductance c which is 1 on E and 0 elsewhere. This defines a Dirichlet form on F_0 by (1.1).

The symmetry group $\mathfrak{G}$ splits the edges of $\Gamma(1)$ into orbits $E_1, \dots, E_l$. Each orbit E_i defines a graph Γ_i, a conductance c_i, and a Dirichlet form $\mathscr{A}_i$ with operator A_i. It suffices to prove that the $A_1, \dots, A_l$ commute, since $\mathbb{D}$ is spanned by $\mathscr{A}_1, \dots, \mathscr{A}_l$. We remark that $c_i(x) \equiv c_i \in \mathbb{R}$ because Γ_i is

$\mathfrak{G}$-invariant and $\mathfrak{G}$ acts transitively on F_0. Let I denote the identity map on $\mathscr{D}_0$, and set $B_i := c_i I - A_i$ for all $1 \leqslant i \leqslant l$. Then the (A_i) commute if and only if the (B_i) do. The (B_i) are the adjacency matrices of the (Γ_i).

For all $x, y \in F_0$ and $1 \leqslant i \leqslant l$,

$$\langle B_i B_j \delta_x, \delta_y \rangle = \sum_{z \in F_0} \langle B_j \delta_x, \delta_z \rangle \langle B_i \delta_z, \delta_y \rangle .$$

The latter sum can be interpreted as the number of paths of length 2 from x to y in $\Gamma(1)$ with the first step along Γ_j and the second along Γ_i. The reflection $\rho_{x,y}$ maps those paths one to one onto the paths of length 2 from y to x in $\Gamma(1)$ with the first step along Γ_j and the second along Γ_i. Hence,

$$\langle B_i B_j \delta_x, \delta_y \rangle = \sum_{z \in F_0} \langle B_j \delta_y, \delta_z \rangle \langle B_i \delta_z, \delta_x \rangle .$$

Since B_i and B_j are symmetric, we arrive at

$$\begin{aligned} \langle B_i B_j \delta_x, \delta_y \rangle &= \sum_{z \in F_0} \langle B_i \delta_x, \delta_z \rangle \langle B_j \delta_z, \delta_y \rangle \\ &= \langle B_j B_i \delta_x, \delta_y \rangle . \end{aligned}$$

Next, we prove the result about the geometry of $\mathbb{P}$. Let $f_1, \dots, f_k$ be the eigenfunctions arising from the simultaneous diagonalization of all elements of $\mathbb{B}$, which exist by the above commutation result [12, Theorem 1.3.19]. Then the spectral representation

$$\begin{aligned} \Sigma : \mathbb{B} &\to \mathbb{R}^k, \\ \mathscr{A} &\mapsto (\mathscr{A}(f_1, f_1), \dots, \mathscr{A}(f_k, f_k)) \end{aligned}$$

is one to one and linear. In particular, $\Sigma(\mathbb{P}) = \mathbb{R}^k_+ \cap \Sigma \circ \Xi(\mathbb{R}^{\dim \mathbb{B}})$. Since $\mathbb{R}^k_+$ is polyhedral and $\dim \Sigma \circ \Xi(\mathbb{R}^{\dim \mathbb{B}}) = \dim \mathbb{B}$, $\Sigma(\mathbb{P})$ is a polyhedral cone in $\mathbb{R}^k$ of dimension $\dim \mathbb{B}$. Hence, $\mathbb{P}$ is simplicial in $\mathbb{B}$. ■

In the latter proof it is sufficient to assume that $\mathfrak{G}$ acts transitively on F_0 instead of generating $\mathfrak{G}$ by $\{ \rho_{x,y} \mid x, y \in F_0, x \neq y \}$.

3. POTENTIAL THEORETIC REMARKS

In this section we will try to explain why the shorted operator is so easily expressed in potential theoretic terms and what this has to do with the concept of effective resistance from electrical network theory. It is used very often in connection with shorted operators, because of its invariance under Φ, the elimination of vertices in an electrical network. A shorted operator can be expressed in terms of the "parallel sum" of operators and vice versa [1, Theorem 4; 3, Theorem 13]. The latter concept was studied potential theoretically in [8].

The Schur complement formula in Section 1 can be modified. By [24, Lemma 2.2.4(iii)] the Moore-Penrose generalized inverse can be replaced by any generalized inverse. On the other hand, for an invertible E_1 the inversion formula for partitioned matrices, [12, p. 18], and (1.2) imply $\Phi(E_1)^{-1} = \pi E_1^{-1}\pi^*$. So there might be some generalized inverse $(\cdot)^-$ such that

$$\Phi(\mathscr{B})^- = \pi\mathscr{B}^-\pi^* \qquad \text{for all} \quad \mathscr{B} \in \mathbb{P}^\circ. \tag{3.1}$$

In potential theory a very prominent generalized inverse of a Dirichlet operator is its *Green's function*. Let us assume that $\mathscr{B}$ is a Dirichlet form on F_1 whose kernel consists only of constants. We choose a reference point $x_1 \in F_0$ and restrict $\mathscr{B}$ to the space of all real valued functions on F_1 that vanish in x_1. Let B denote the Dirichlet operator of the old form and $\pi_{\{x_1\}}$ the orthogonal projection onto functions that vanish in x_1; then $\pi_{\{x_1\}}B\pi^*_{\{x_1\}}$ is the Dirichlet operator of the new form, and

$$G := \left(\pi_{\{x_1\}}B\pi^*_{\{x_1\}}\right)^{-1}$$

is called the Green's function of B on $F_1 \setminus \{x_1\}$. By the standard link between potential theory and symmetric Markov processes, the Green's function can be interpreted in terms of expected local times [7, Formula 7]. The generalized inverse G indeed fulfils (3.1) [20, Proposition 3.3].

Another generalized inverse frequently used in connection with shorted operators is the effective resistance $R(x, y)$ between two different vertices x and y of an electrical network. It is also known as the "Campbell-Youla inverse" [27, Formula 33]. Its popularity is due to the fact that it also fulfils (3.1), as we will see. Following Doyle and Snell [6, p. 62], we define the effective resistance via the Dirichlet form. Let $x, y \in F_0$ with $x \neq y$, and define $e_{x,y}$ by $e_{x,y}(x) = 1$, $e_{x,y}(y) = 0$, and $e_{x,y}$ harmonic with respect to

B on $F_1 \setminus \{x, y\}$. Such a function is called an "equilibrium potential" in potential theory. Now

$$R(x,y)^{-1} := \mathscr{B}(e_{x,y}, e_{x,y}).$$

Since $e_{x,y}$'is the solution of a certain Dirichlet problem, the Dirichlet principle tells us that

$$R(x,y)^{-1} = \inf\left\{ \frac{\mathscr{B}(f,f)}{[f(x) - f(y)]^2} : f \in \mathscr{D}_1,\ f(x) \neq f(y) \right\}.$$

By definition we have $B(G\delta_x) = \delta_x - \delta_{x_1}$. Hence $G(\delta_x - \delta_y)$ is also harmonic on $F_1 \setminus \{x, y\}$ with respect to $\mathscr{B}$. Thus,

$$R(x,y)^{-1} = \frac{\mathscr{B}\big(G(\delta_x - \delta_y), G(\delta_x - \delta_y)\big)}{\big[G(\delta_x - \delta_y)(y) - G(\delta_x - \delta_y)(x)\big]^2}$$

$$= \big\langle G(\delta_x - \delta_y), \delta_x - \delta_y \big\rangle_1^{-1}.$$

We finally arrive at

$$R(x,y) = \big\langle G(\delta_x - \delta_y), \delta_x - \delta_y \big\rangle_1. \tag{3.2}$$

In particular, $R(x, x_1) = \langle G\delta_x, \delta_x \rangle_1$. The norm defined by G is known in potential theory as "Maeda's energy norm" [18, Corollary 4.5, Theorem 4.2]. So the effective resistance between x and y is nothing else than Maeda's energy norm of $\delta_x - \delta_y$. The multiple ν of $\delta_x - \delta_y$ with $G\nu = e_{x,y}$ is called an "equilibrium measure." On the other hand we can reproduce G from R by the second polarization identity.

$$\langle G\delta_x, \delta_y \rangle_1 = \tfrac{1}{2}\big[R(x, x_1) + R(y, x_1) - R(x,y)\big].$$

So R is just another way to write down G, and vice versa. Since (3.1) holds for G, it also holds for R.

The Green's function G defines a scalar product. In particular,

$$d(x,y) := R(x,y)^{1/2}, \qquad d(x,x) := 0,$$

for all $x, y \in F_1$, defines a metric on F_1, the so called effective resistance metric [14, Definition 0.5]. In the light of (3.2) it is a distance arising from an inner product. In this sense a Dirichlet form defines a geometry on a graph. We remark that R does not depend on x_1, although G does.

We have seen two ways to deal with effective resistances in potential theory: as a Dirichlet norm of equilibrium potentials and as the energy norm of equilibrium measures. A third, more general way is to view effective resistance as a "relative capacity." Consider an $\mathscr{A} \in \mathbb{D} \cap \mathbb{P}^\circ$ and $x, y \in F_0$ with $x \neq y$. Let $\mathscr{A}_{F_0 \setminus \{x\}}$ denote the restriction of $\mathscr{A}$ to $\mathscr{D}_{F_0 \setminus \{x\}} = \{f \in \mathscr{D}_0 \mid f(x) = 0\}$. Now define

$$\mathrm{Cap}_{F_0 \setminus \{x\}}(y) := \inf\left\{\mathscr{A}_{F_0 \setminus \{x\}}(f, f) \,\middle|\, f \in \mathscr{D}_{F_0 \setminus \{x\}},\ f(y) \geqslant 1\right\}.$$

Then

$$\mathrm{Cap}_{F_0 \setminus \{x\}}(y) = R(x, y)^{-1}.$$

We name $\mathrm{Cap}_{F_0 \setminus \{x\}}(y)$ the capacity of y relative to x.

All three concepts—Dirichlet forms, energy forms, and capacities—are central in potential theory. They provide three different ways to understand the shorted operator and/or effective resistance. The interpretations as electrical networks, Dirichlet forms, and distances are by no means new. They correspond for example to the models C, B, and D in [9]. The main potential theoretic message therefore is the dominant role of Green's functions and energy forms together with their standard probabilistic interpretations.

4. SHORTING INFINITE MODELS

Graphically $\Lambda^n(\mathscr{A})$ can be interpreted in the following way: Refine the initial "grid" F_0 n-fold, that is $F_n := \Psi^n(F_0)$. Then refine the Dirichlet form $\mathscr{A} \in \mathbb{D}$ n-fold, that is, $\mathscr{A}_n := \Psi^n(\mathscr{A})$. Now eliminate all vertices of $F_n \setminus F_0$ from F_n by

$$\Lambda^n(\mathscr{A})(f, f) = \inf\left\{\mathscr{A}_n(g, g) \,\middle|\, g : F_n \to \mathbb{R},\ g|_{F_0} = f\right\},$$

as in [20, Formula 6.4]. In this sense the renormalization limit

$$\lim_{n \to \infty} (\Lambda/\gamma)^n(\mathscr{A}) =: \mathscr{A}_\infty \tag{4.1}$$

is the result of eliminating all vertices but those in F_0 from an infinite "grid" $F := \bigcup_{n \in \mathbb{N}} F_n$. It is a solution of the eigenvalue problem $\Lambda(\mathscr{A}_\infty) = \gamma \mathscr{A}_\infty$.

More precisely, we define a Dirichlet form $(\mathscr{J}, \mathscr{D})$ on $L^2(X, \mu)$ using the fact that F is dense in X. We set $\mathscr{J}_0 := \mathscr{A}_\infty$ and

$$\begin{aligned} \mathscr{D} &:= \left\{ f \in \mathscr{C}(X, \mathbb{R}) \,\middle|\, \sup_{n \in \mathbb{N}} \gamma^{-n} \mathscr{J}_n(f|_{F_n}, f|_{F_n}) < \infty \right\}, \\ \mathscr{J}(f,f) &:= \lim_{n \to \infty} \gamma^{-n} \mathscr{J}_n(f|_{F_n}, f|_{F_n}). \end{aligned} \tag{4.2}$$

This is a local, regular Dirichlet form [16, Theorem 4.14]. In particular, for $f \in \mathscr{D}_0$ we have

$$\mathscr{A}_\infty = \inf\{\mathscr{J}(g, g) \mid g \in \mathscr{D},\ g|_{F_0} = f\}.$$

This equality coincides with the above shorting of an infinite "grid." For $f : F_n \to \mathbb{R}$, we even have

$$\gamma^{-n} \mathscr{J}_n(f, f) = \mathscr{J}(H_{X \setminus F_n} f, H_{X \setminus F_n} f), \tag{4.3}$$

where $H_{X \setminus F_n} f$ is a function harmonic on $X \setminus F_n$ with respect to $\mathscr{J}$ and boundary data f on F_n. This is due to the fact that $\mathscr{A}_\infty$ is an eigenvector of Λ. For $n = 0, 1$ we recover our eigenvalue equation.

The result (4.1) corresponds to a "homogenization" property in the sense of [15]: We start with $\mathscr{J}_0' := \mathscr{A}$ and try the construction (4.2). This time we get the "Γ-convergence" of $(\mathscr{J}_n')_{n \in \mathbb{N}}$ to the limit $\mathscr{J}' = \mathscr{J}$. But we lose the primed version of (4.3). Nevertheless the original version of (4.3) still holds. In this sense the limiting form $\mathscr{J}$ is more homogeneous than the approximating forms $(\mathscr{J}_n')_{n \in \mathbb{N}}$, that is, homogenization took place.

The concepts of Dirichlet operator, Green's function, effective resistance, Maeda's energy norm, and relative capacity can all be formulated for the continuous model as well, and their interrelations remain the same with a little more notational precaution. In particular the connection between effective resistance and the energy norm remains valid. Again an effective resistance defines an energy norm, which defines a Dirichlet form. In transient cases, points have to be replaced by nonpolar sets [18, 25].

REFERENCES

1 W. N. Anderson, Jr., Shorted operators, *SIAM J. Appl. Math.* 20:520–525 (1971).

2 W. N. Anderson, Jr., and G. E. Trapp, Shorted operators, II, *SIAM J. Appl. Math.* 28:60–71 (1975).

3 W. N. Anderson, Jr., and G. E. Trapp, Matrix operations induced by electrical network connections—a survey, in *Constructive Approaches to Mathematical Models* (C. V. Coffman and G. J. Fix, Eds.), 1979, pp. 53–73. Academic.

4 M. T. Barlow, Random walks and diffusions on fractals, in *Proceedings of the International Congress of Mathematicians Kyoto 1990*, Springer-Verlag, Tokyo, 1991, pp. 1025–1035.

5 D. Carlson, What are Schur complements, anyway? *Linear Algebra Appl.* 74:257–275 (1986).

6 P. G. Doyle and J. L. Snell, *Random Walks and Electric Networks*, Carus Math. Monographs 22, Math. Assoc. Amer., Washington, 1984.

7 E. B. Dynkin and A. A. Jushkevich, *Markov Processes: Theorems and Problems*, Plenum, New York, 1969.

8 S.-L. Eriksson and H. Leutwiler, A potential-theoretic approach to parallel addition, *Math. Ann.* 274:301–317 (1986).

9 M. Fiedler, Aggregation in graphs, in *Combinatorics*, Colloq. Math. Soc. János Bolyai 18, Keszthely, 1976, pp. 315–330.

10 M. Fukushima, Y. Oshima, and M. Takeda, *Dirichlet Forms and Symmetric Markov Processes*, de Gruyter Stud. Math. 19, Walter deGruyter, New York, 1994.

11 K. Hattori, T. Hattori, and H. Watanabe, Gaussian field theories on general networks and the spectral dimension, *Progr. Theoret. Phys. Suppl.* 92:108–143 (1987).

12 R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge U.P., 1985.

13 R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*, Cambridge U.P., 1991.

14 J. Kigami, Harmonic calculus on limits of networks and its application to dendrites, *J. Funct. Anal.* 128:48–68 (1995).

15 S. M. Kozlov, Harmonization and homogenization on fractals, *Commun. Math. Phys.* 153:339–357 (1993).

16 S. Kusuoka, Diffusion processes on nested fractals, in *Statistical Mechanics and Fractals* (R. L. Dobrushin and S. Kusuoka, Eds.), Lecture Notes in Math. 1567, Springer-Verlag, Berlin, 1993, pp. 39–98.

17 T. Lindstrøm, *Brownian Motion on Nested Fractals*, Mem. Amer. Math. Soc. 83(420), Amer. Math. Soc., Providence, 1990.

18 F.-Y. Maeda, *Dirichlet Integrals on Harmonic Spaces*, Lecture Notes in Math. 803, Springer-Verlag, Berlin, 1980.

19 V. Metz, Hilbert's projective metric on cones of Dirichlet forms, *J. Funct. Anal.* 127:438–455 (1995).

20 V. Metz, Renormalization of finitely ramified fractals, *Proc. Roy. Soc. Edinburgh* 125A:1085–1104 (1995).

21 V. Metz, Renormalization on fractals, in *Potential Theory-ICPT94, Proceedings of the International Conference on Potential Theory 1994, Kouty*, (J. Král et al., Eds.), deGruyter, Berlin-New York, 1996, pp. 413–422.

22 V. Metz, Renormalization contracts on nested fractals, *J. Reine Angew. Math.*, 480:161–175 (1996).

23 R. D. Nussbaum, *Hilbert's Projective Metric and Iterated Nonlinear Maps*, Mem. Amer. Math. Soc. 75(391), Amer. Math. Soc., Providence, 1988.

24 C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and Its Applications*, Wiley, 1971.

25 M. Röckner, Self-adjoint harmonic spaces and Dirichlet forms, *Hiroshima Math. J.* 14:55–66 (1984).

26 C. Sabot, Existence et Unicité de la Diffusion sur un Fractal, Dissertation, Univ. Pierre et Marie Curie, Lab. de Probabilité, Paris, 1995.

27 G. E. Sharpe and G. P. H. Styan, Circuit duality and the general network inverse, *IEEE Trans. Circuit Theory* CT-12:22–27 (1965).

Received 11 November 1995; final manuscript accepted 6 May 1996