Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 1, No. 3, pp. 509–526, September 2005

# SVM KERNEL BY ELECTRIC NETWORK

Hiroshi Hirai, Kazuo Murota and Masaki Rikitoku

*Dedicated to Professor Toshihide Ibaraki on his 65th birthday.*

**Abstract:** This paper investigates support vector machine (SVM) with a discrete kernel, named electric network kernel, defined on the vertex set of an undirected graph. Emphasis is laid on mathematical analysis of its theoretical properties with the aid of electric network theory and the theory of discrete metrics. SVM with this kernel admits physical interpretations in terms of resistive electric networks; in particular, the SVM decision function corresponds to an electric potential. Preliminary computational results indicate reasonable promise of the proposed kernel in comparison with the Hamming and diffusion kernels.

**Key words:** *support vector machine (SVM), discrete kernel, discrete Green's function, tree metric, electric network, inverse M-matrix*

**Mathematics Subject Classification:** *90C90, 68T05*

## 1 Introduction

Support vector machine (SVM) has come to be very popular in machine learning and data mining communities. SVM is a binary classifier using an optimal hyperplane learned from given training data. Through *kernel functions*, which are a kind of similarity functions defined on the data space, the data can be implicitly embedded into a high (possibly infinite) dimensional Hilbert space. With this *kernel trick*, SVM achieves a nonlinear classification with low computational cost.

Input data from real world problems, such as text data, DNA sequences and hyperlinks in World Wide Web, is often endowed with discrete structures. Theory and application of "kernels on discrete structures" are pioneered by D. Haussler [10], C. Watkins [22] and R. I. Kondor and J. Lafferty [13]. Haussler and Watkins independently introduced the concept of *convolution kernels*. Kondor and Lafferty utilized spectral graph theory to introduce *diffusion kernels*, which are discrete kernels defined on vertices of graphs.

In this paper we propose a novel class of discrete kernels on vertices of an undirected graph. Our approach is closely related to that of Kondor and Lafferty, but is based on electric network theory rather than on spectral graph theory. Accordingly we will name the proposed kernels *electric network kernels*. SVM using an electric network kernel admits natural physical interpretations. The vertices with positive label and negative label correspond, respectively, to terminals with $+1$ electric potential and $-1$ electric potential. The resulting decision function corresponds to an electric potential, and the separating hyperplane to points with potential equal to zero.

Emphasis is laid on mathematical analysis of the electric network kernel with the aid of electric network theory and the theory of discrete metrics. An interesting link to discrete metrics is revealed by considering the special case where the underlying graph is a tree. Then the electric network kernel is equivalent, in a nontrivial sense, to a *tree metric*, which is a fundamental concept in phylogeny [18]. Combination of this observation with the Gomory-Hu cut tree known in network flow theory (see, e.g., [5]) naturally leads to a discrete kernel based on the minimum cuts in an undirected graph. Another interesting special case is where the underlying graph is a hypercube. By exploiting symmetry of a hypercube, we provide an explicit formula for the electric network kernel, which makes it possible to apply the electric network kernel to large-scale practical problems. In our preliminary computational experiment the electric network kernel shows fairly good performance for some data sets, as compared with the Hamming and diffusion kernels.

This paper is organized as follows. In Section 2, we review SVM and its formulation as optimization problems. In Section 3, we propose our kernel and investigate its properties. Physical interpretations to SVM with our kernel are also explained. In Section 4, we consider the case of a tree and indicate links to a tree metric. In Section 5, we deal with the case of a hypercube, and show some computational results for some real world problems.

## 2 Support Vector Machines

In this section, we review SVM and its formulation as optimization problems; see [17], [21] for details. Let $\mathcal{X}$ be an input data space, e.g. $\mathbf{R}^n$, $\{0,1\}^n$, text data and DNA sequence, etc. A symmetric function $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ is said to be a *kernel* on $\mathcal{X}$ if it satisfies the *Mercer condition*:

$$\begin{aligned}&\text{For any finite subset } Y \text{ of } \mathcal{X}\\ &\text{matrix } (K(x,y) \mid x,y \in Y) \text{ is positive semidefinite.}\end{aligned} \tag{2.1}$$

For a kernel $K$, it is well known that there exists some Hilbert space $\mathcal{H}$ with inner product $\langle \cdot,\cdot \rangle$ and a map $\phi : \mathcal{X} \to \mathcal{H}$ such that

$$K(x,y) = \langle \phi(x), \phi(y) \rangle \quad (x,y \in \mathcal{X}).$$

Given a labeled training set $(x_1,\eta_1),(x_2,\eta_2),\dots,(x_m,\eta_m) \in \mathcal{X} \times \{\pm 1\}$, SVM classifier is obtained by solving the optimization problem

$$\begin{aligned}\min_{\alpha \in \mathbf{R}^m} \quad & \frac{1}{2}\sum_{1\le i,j\le m} \alpha_i\alpha_j\eta_i\eta_j K(x_i,x_j) - \sum_{1\le i\le m}\alpha_i\\ \text{s.t.} \quad & \sum_{1\le i\le m}\eta_i\alpha_i = 0,\ 0 \le \alpha_i \le C \quad (i=1,\dots,m),\end{aligned}$$

where $C$ is a penalty parameter that is a positive real number or $+\infty$. If $C = +\infty$, it is called the *hard margin SVM* formulation. If $C < +\infty$, it is called the *1-norm soft margin SVM* formulation.

For our purpose, it is convenient to consider the equivalent problem

$$\begin{aligned}[\text{SVM}] : \min_{u \in \mathbf{R}^m} \quad & \frac{1}{2}\sum_{1\le i,j\le m} u_iu_jK(x_i,x_j) - \sum_{1\le i\le m}\eta_iu_i\\ \text{s.t.} \quad & \sum_{1\le i\le m} u_i = 0, && (2.2)\\ & 0 \le \eta_iu_i \le C \quad (i=1,\dots,m), && (2.3)\end{aligned}$$

where $u_i = \eta_i \alpha_i$ for $i = 1, \ldots, m$.

Let $u^* \in \mathbf{R}^m$ be an optimal solution of the problem [SVM] and $b^* \in \mathbf{R}$ be the Lagrange multiplier of constraint (2.2) at $u^*$, where the Lagrange function of [SVM] is supposed to be defined as

$$\begin{aligned} L(u, \lambda, \mu, b) &= \frac{1}{2} \sum_{1 \leq i,j \leq m} u_i u_j K(x_i, x_j) - \sum_{1 \leq i \leq m} \eta_i u_i \\ &\quad - \sum_{1 \leq i \leq m} \lambda_i \eta_i u_i - \sum_{1 \leq i \leq m} \mu_i (\eta_i u_i - C) + b \sum_{1 \leq i \leq m} u_i, \end{aligned}$$

where $u \in \mathbf{R}^m$, $\lambda, \mu \in \mathbf{R}^m_{\geq 0}$, and $b \in \mathbf{R}$. Then the decision function $f : \mathcal{X} \to \mathbf{R}$ is given as

$$f(x) = \sum_{i=1}^{m} u_i^* K(x_i, x) + b^* \quad (x \in \mathcal{X}). \tag{2.4}$$

That is, we classify a given data $x$ according to the sign of $f(x)$. A data $x_i$ with $\eta_i u_i^* > 0$ is called a *support vector*. In the case of the 1-norm soft margin SVM, a support vector $x_i$ is called a *normal support vector* if $0 < \eta_i u_i^* < C$ and a *bounded support vector* if $\eta_i u_i^* = C$.

## 3 Proposed Kernel and Its Properties

Let $(V, E, r)$ be a resistive electric network with vertex set $V$, edge set $E$, and Ohmic resistors on edges with the resistances represented by $r : E \to \mathbf{R}_{>0}$. We assume that the graph $(V, E)$ is connected. Let $D : V \times V \to \mathbf{R}$ be a function on $V$ defined as

$$D(x, y) = \text{resistance between } x \text{ and } y \quad (x, y \in V). \tag{3.1}$$

Clearly, $D$ is a nonnegative symmetric function with zero diagonals, i.e., $D$ is a *distance function* on $V$. Fix some vertex $x_0 \in V$ as a root, and define a symmetric function $K : V \times V \to \mathbf{R}$ on $V$ as

$$K(x, y) = \{D(x, x_0) + D(y, x_0) - D(x, y)\}/2 \quad (x, y \in V). \tag{3.2}$$

Seeing that $K(x_0, y) = 0$ for all $y \in V$, we define a symmetric matrix $\hat{K}$ by

$$\hat{K} = (K(x, y) \mid x, y \in V \setminus \{x_0\}). \tag{3.3}$$

**Remark 3.1.** Given a distance function $D$, the function $K$ defined by (3.2) is called the *Gromov product.*

Let $L$ be the *node admittance matrix* defined as

$$L(x, y) = \begin{cases} \sum\{(r(e))^{-1} \mid x \text{ is an endpoint of } e \in E\} & \text{if } x = y, \\ -(r(xy))^{-1} & \text{if } xy \in E, \\ 0 & \text{otherwise,} \end{cases} \quad (x, y \in V). \tag{3.4}$$

If all resistances are equal to 1, then $L$ coincides with the *Laplacian matrix* of graph $(V, E)$. Let $\hat{L}$ be a symmetric matrix defined as

$$\hat{L} = (L(x, y) \mid x, y \in V \setminus \{x_0\}).$$

Note that $\hat{L}$ satisfies

$$\hat{L}(x,y) \leq 0 \quad (x \neq y), \tag{3.5}$$
$$\sum_{z\in V\setminus\{x_0\}} \hat{L}(x,z) \geq 0 \quad (x \in V \setminus \{x_0\}). \tag{3.6}$$

A matrix is called an *M-matrix* if it can be represented as $sI-B$ with a nonnegative matrix $B$ and a real number $s$ not smaller than the spectral radius of $B$. It is known [1] that symmetric matrices satisfying the off-diagonal nonpositivity (3.5) and the diagonal dominance (3.6) are exactly the same as diagonally dominant symmetric M-matrices. Hence $\hat{L}$ is a diagonally dominant symmetric M-matrix. Since the graph $G$ is connected and $\hat{L}$ is a submatrix of $L$ obtained by the deletion of the column and row corresponding $x_0$, $\hat{L}$ is nonsingular and positive definite. A matrix whose inverse is an M-matrix is called an *inverse M-matrix*. The following relationship between $K$ and $L$ is well known in electric network theory; see [7] for example.

**Proposition 3.2.** *We have* $\hat{K}^{-1} = \hat{L}$ *and*

$$\begin{cases} LK\xi = \xi & (\xi \in \mathbf{R}^V,\ \sum_{x\in V}\xi(x) = 0), \\ KLp = p & (p \in \mathbf{R}^V,\ p(x_0) = 0). \end{cases} \tag{3.7}$$

*In particular* $\hat{K}$ *is an inverse M-matrix.*

*Proof.* A proof is provided for completeness. For $x, y \in V$, the resistance between $x$ and $y$ is given by the electric potential difference between $x$ and $y$ when unit electric current flows from $x$ to $y$. By Ohm's law, the electric potential $p : V \to \mathbf{R}$ is given by the solution of linear equation

$$Lp = \chi_x - \chi_y, \tag{3.8}$$

where $\chi_x$ is the unit vector defined as $\chi_x(z) = 1$ if $z = x$ and 0 otherwise. With an additional condition $p(x_0) = 0$, then the solution of (3.8) is uniquely determined. Hence the resistance $D(x,y)$ is given as

$$\begin{aligned} D(x,y) &= p(x) - p(y) \\ &= \begin{cases} \hat{L}^{-1}(x,x) + \hat{L}^{-1}(y,y) - 2\hat{L}^{-1}(x,y) & \text{if } x, y \in V \setminus \{x_0\}, \\ \hat{L}^{-1}(x,x) & \text{if } x \in V \setminus \{x_0\}, y = x_0, \\ \hat{L}^{-1}(y,y) & \text{if } y \in V \setminus \{x_0\}, x = x_0. \end{cases} \end{aligned}$$

From this, we obtain

$$K(x,y) = \{D(x,x_0) + D(y,x_0) - D(x,y)\}/2 = \hat{L}^{-1}(x,y)$$

for $x, y \in V \setminus \{x_0\}$. Equations in (3.7) follow from the facts that each column sum of $L$ equals zero, $K(\cdot, x_0) = K(x_0, \cdot) = 0$ and that $\hat{K}^{-1} = \hat{L}$. □

Hence, $K$ in (3.2) is positive semidefinte and satisfies the Mercer condition. We shall call such $K$ an *electric network kernel*.

**Remark 3.3.** Chung and Yau [4] considered *discrete Green's function* on a graph $(V, E, r)$, which is a function $G : V \times V \to \mathbf{R}$ satisfying

$$GLp = p \quad (p \in \mathbf{R}^V,\ p(x) = \sigma(x)\ (x \in \delta V)),$$

where $\sigma : \delta V \to \mathbf{R}$ with $\delta V \subseteq V$ represents a prespecified boundary condition. Equation (3.7) implies that our electric network kernel $K$ of $(V, E, r)$ with root $x_0$ coincides with this discrete Green's function with $\delta V = \{x_0\}$ and $\sigma(x_0) = 0$.

We consider the SVM on electric network $(V, E, r)$ with the kernel $K$ of (3.2). Let $\{(x_i, \eta_i)\}_{i=1,\dots,m} \subseteq V \times \{\pm 1\}$ be a training data set, where we assume that $x_i$ $(i = 1, \dots, m)$ are all distinct. Just as the SVM with a diffusion kernel, we assume that $\{x_1, \dots, x_m\}$ is a subset of the vertex set $V$; accordingly we put $V = \{x_1, \dots, x_n\}$ with $n \geq m$.

**Lemma 3.4.** *The optimization problem* [SVM] *is determined independently of the choice of a root* $x_0 \in V$.

*Proof.* The objective function of [SVM] is in fact independent of $x_0$, since its quadratic term can be rewritten as

$$\begin{aligned}
\textstyle\sum_{i,j} u_i u_j K(x_i, x_j) &= \textstyle\sum_{i,j} u_i u_j (D(x_i, x_0) + D(x_j, x_0) - D(x_i, x_j))/2 \\
&= \textstyle\sum_j u_j \sum_i u_i D(x_i, x_0) - (1/2) \sum_{i,j} u_i u_j D(x_i, x_j) \\
&= -(1/2) \textstyle\sum_{i,j} u_i u_j D(x_i, x_j),
\end{aligned}$$

where the last equality follows from the constraint (2.2). □

Next we give physical interpretations to the problem [SVM] with the aid of nonlinear network theory (see [12, Chapter IV]). Suppose that we are given an electric network $(V, E, r)$ and labeled training data set $\{(x_i, \eta_i)\}_{i=1,\dots,m} \subseteq V \times \{\pm 1\}$, where $x_1, \dots, x_m$ are all distinct. We connect voltage sources to $(V, E, r)$ as follows:

> For each $x_i$ with $1 \leq i \leq m$, connect to the earth a voltage source whose electric potential is $\eta_i$ and the current flowing into $x_i$ is restricted to $[0, C]$ if $\eta_i = 1$ and $[-C, 0]$ if $\eta_i = -1$.

By using voltage sources, current sources and diodes, this network can be realized as in Figure 1.

Let $A = (A(x, e) \mid x \in V, e \in E)$ be the incidence matrix of $(V, E)$ with some fixed orientation of edges and let $R = \mathrm{diag}(r(e) \mid e \in E)$ be the diagonal matrix whose diagonals are the resistances of edges.

The electric current in this network is given as an optimal solution of the problem:

$$\begin{aligned}
[\text{FLOW}] : \min_{(\zeta, \xi)} \quad & \frac{1}{2} \zeta^\top R \zeta - \sum_{i=1}^{m} \eta_i \xi_i \\
\text{s.t.} \quad & A\zeta = \begin{pmatrix} \xi \\ 0 \end{pmatrix}, \\
& \sum_{1 \leq i \leq m} \xi_i = 0, \quad 0 \leq \eta_i \xi_i \leq C \ (i = 1, \dots, m),
\end{aligned}$$

where $\zeta$ represents the currents in edges and $\xi_i$ represents the current flowing into $x_i$ for $i = 1, \dots, m$. The first and second terms of the objective function of [FLOW] represents current potential of edges $E$ and of the voltage sources respectively. The electric potential of this network is given as an optimal solution of the problem:

$$[\text{POT}] : \min_{p \in \mathbf{R}^n} \quad \frac{1}{2} p^\top A R^{-1} A^\top p + C \sum_{i=1}^{m} \max\{0, 1 - \eta_i p_i\},$$

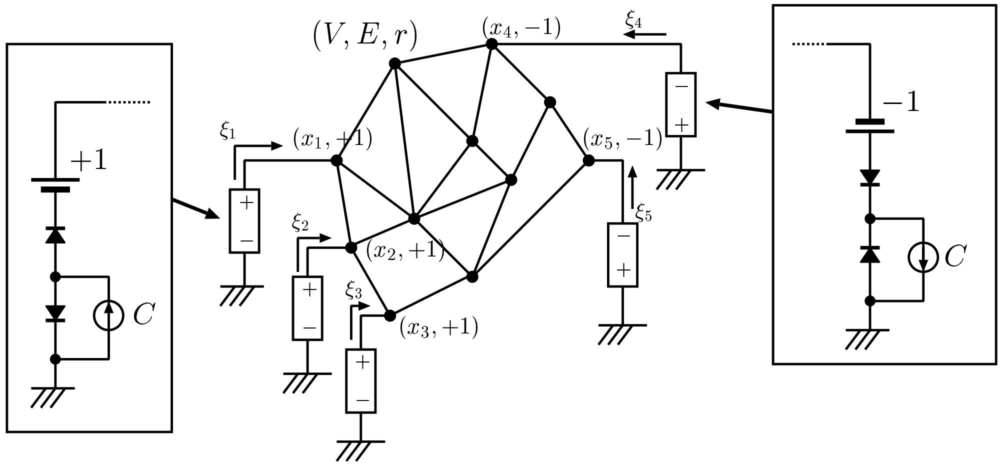


Figure 1: Physical interpretation

where $p_i$ represents the potential at vertex $x_i$ for $i = 1, \ldots, n$. The first and second terms of the objective function of [POT] represent voltage potentials of edges $E$ and of the voltage sources respectively. In fact, [FLOW] and [POT] are a dual pair. The optimality condition is given as follows. Feasible $(\zeta, \xi)$ and $p$ are optimal solutions of [FLOW] and [POT] if and only if they satisfy

$$\forall x_i x_j \in E: \ p_i - p_j = \zeta(x_i x_j) R(x_i x_j) \text{ (Ohm's law)}, \tag{3.9}$$

$$\forall i \in \{1, \ldots, m\}: \begin{cases} 0 < \eta_i \xi_i < C & \Rightarrow \quad \eta_i p_i = 1, \\ \eta_i \xi_i = 0 & \Rightarrow \quad \eta_i p_i \geq 1, \\ \eta_i \xi_i = C & \Rightarrow \quad \eta_i p_i \leq 1. \end{cases} \tag{3.10}$$

**Proposition 3.5.** *The electric current $(\zeta^*, \xi^*)$ in this network is uniquely determined. If there exists $i \in \{1, \ldots, m\}$ with $0 < \eta_i \xi_i^* < C$, then the electric potential is also uniquely determined.*

*Proof.* The first assertion follows from the uniqueness theorem [12, Theorem 16.2]. If such $\xi_i^*$ exists, from complementarity condition (3.10), any optimal solution $p^*$ of [POT] must satisfy $p_i^* = \eta_i$. Consequently, the potentials of other vertices are also uniquely determined by Ohm's law (3.9). □

The following theorem indicates the relationship between SVM problem and this electric network.

**Theorem 3.6.** *Let $u^*$ be the optimal solution of* [SVM]. *Then $u_i^*$ coincides with the electric current flowing into $x_i$ for $i = 1, \ldots, m$. Moreover, the decision function $f$ of* (2.4) *for* [SVM] *is an electric potential.*

*Proof.* The problem [FLOW] is equivalent to

$$\begin{aligned} [\text{FLOW}'] : \min_{\xi} \quad & W(\xi) - \sum_{i=1}^{m} \eta_i \xi_i \\ \text{s.t.} \quad & \sum_{1 \leq i \leq m} \xi_i = 0, \quad 0 \leq \eta_i \xi_i \leq C \ (i = 1, \ldots, m), \end{aligned}$$

where $W : \mathbf{R}^m \to \mathbf{R}$ is defined as

$$W(\xi) = \min_{\zeta} \left\{ \frac{1}{2}\zeta^\top R\zeta \;\middle|\; A\zeta = \begin{pmatrix} \xi \\ 0 \end{pmatrix} \right\}.$$

By the Lagrange multiplier method, the minimizer $\zeta^*$ satisfies

$$R\zeta^* - A^\top \mu^* = 0,\ A\zeta^* - \begin{pmatrix} \xi \\ 0 \end{pmatrix} = 0,$$

where $\mu^*$ is the optimal Lagrange multiplier with $\mu^*(x_0) = 0$. From this, we have

$$W(\xi) = \frac{1}{2}\mu^{*\top} L\mu^*,\ L\mu^* = \begin{pmatrix} \xi \\ 0 \end{pmatrix},$$

where $L = AR^{-1}A^\top$ is the node admittance matrix. We define $K$ by (3.2). From the second equation above and (3.7), we obtain

$$K \begin{pmatrix} \xi \\ 0 \end{pmatrix} = KL\mu^* = \mu^*.$$

Hence we have the desired formula

$$W(\xi) = \frac{1}{2}\mu^{*\top} L\mu^* = \frac{1}{2}(\xi^\top 0^\top)KLK \begin{pmatrix} \xi \\ 0 \end{pmatrix} = \frac{1}{2} \sum_{1 \le i,j \le m} \xi_i \xi_j K(x_i, x_j),$$

where the third equality follows from (3.7). This implies that the problem [FLOW$'$] coincides with [SVM]. Hence we have $u^* = \xi^*$, where $\xi^*$ denotes the uniquely determined optimal solution of [FLOW$'$].

Next we show the latter half of the claim. Let $f : V \to \mathbf{R}$ be the decision function given by (2.4) with $u^* = \xi^*$, i.e.,

$$f = K \begin{pmatrix} \xi^* \\ 0 \end{pmatrix} + b^*\mathbf{1},$$

where $\mathbf{1}$ is the $n$-dimensional vector with all elements 1. By well-known optimality criterion for the SVM problem [17], $\xi^*$ and $f$ satisfy the condition (3.10) with $p = f$. Let $\zeta^*$ be defined as $\zeta^* = R^{-1}A^\top f$. By this construction, $\zeta^*$ and $f$ satisfy Ohm's law (3.9) with $p = f$. Furthermore, feasibility of $\zeta^*$ is shown as

$$A\zeta^* = AR^{-1}A^\top f = L\left\{K \begin{pmatrix} \xi^* \\ 0 \end{pmatrix} + b^*\mathbf{1}\right\} = \begin{pmatrix} \xi^* \\ 0 \end{pmatrix},$$

where the last equality follows from (3.7) and $L\mathbf{1} = 0$. This implies that $(\zeta^*, \xi^*)$ and $f$ are the optimal solutions of [FLOW] and [POT], respectively. □

From Proposition 3.5 and Theorem 3.6, we see that the electric potential coincides with the decision function of [SVM], provided that the optimal solution of [SVM] has a normal support vector. Furthermore, the Lagrange multiplier $b^*$ corresponds to the electric potential of the root vertex $x_0$, if the potential is normalized in such a way that the earth has zero electric potential. Hence, in the SVM with our electric network kernel, the following correspondence holds.

| SVM | electric network |
|---|---|
| positive label data | $+1$ voltage sources |
| negative label data | $-1$ voltage sources |
| optimal solution of [SVM] | electric current from voltage sources |
| decision function | electric potential |

Next we consider the case of the hard-margin SVM. The following proposition indicates that solving [SVM] with $C = +\infty$ reduces to solving linear equations.

**Proposition 3.7.** *For the electric network kernel $K$, an optimal solution $u^*$ of the equality constrained optimization problem*

$$\begin{array}{lll} [\mathrm{SVM}'] : \displaystyle\min_{u \in \mathbf{R}^m} & \displaystyle\frac{1}{2} \sum_{1 \leq i,j \leq m} u_i u_j K(x_i, x_j) - \sum_{1 \leq i \leq m} \eta_i u_i \\ \qquad\qquad \text{s.t.} & \displaystyle\sum_{1 \leq i \leq m} u_i = 0 \end{array}$$

*is also optimal to* [SVM] *with $C = +\infty$.*

*Proof.* Suppose that $\eta_i = +1$ for $1 \leq i \leq k$ and $\eta_i = -1$ for $k+1 \leq i \leq m$ for $1 \leq k < m$. By a variant of Lemma 3.4, we may take $x_m$ as the root. Then problem [SVM$'$] is equivalent to

$$\min_{u \in \mathbf{R}^{m-1}} \frac{1}{2} \sum_{1 \leq i,j \leq m-1} u_i u_j K(x_i, x_j) - \sum_{1 \leq i \leq k} 2u_i,$$

where we substitute $u_m = -\sum_{1 \leq i \leq m-1} u_i$ in [SVM$'$]. Let $\overline{K} = (K(x_i, x_j) \mid 1 \leq i, j \leq m-1)$. Hence the optimal solution $u^* \in \mathbf{R}^m$ is given by

$$\begin{aligned} u_i^* &= 2 \sum_{1 \leq j \leq k} (\overline{K}^{-1})_{ij} \quad (1 \leq i \leq m-1), \\ u_m^* &= -2 \sum_{1 \leq j \leq k} \sum_{1 \leq h \leq m-1} (\overline{K}^{-1})_{hi}. \end{aligned}$$

Since $\overline{K}$ is an inverse $M$-matrix by Proposition 3.2, we have

$$u_i^* \geq 0 \quad (1 \leq i \leq k), \quad u_i^* \leq 0 \quad (k+1 \leq i \leq m).$$

Hence $u^*$ satisfies the inequality constraint of [SVM] and is optimal. □

From Proposition 3.7, the hard-margin SVM problem can be completely reduced to the problem of linear resistive network without nonlinear devices, e.g., diodes. In particular, we see the following fundamental properties.

- decision function, or equivalently, electric potential $f$ is fixed to be $f(x_i) = \eta_i$ for each data $x_i$.
- for an edge $xy$, $f(x) > f(y)$ if and only if electric current flows from $x$ to $y$ (Ohm's law).
- there is no electric current flowing into (going from) $+1$ $(-1)$ voltage sources.

The following corollaries follow from above arguments of linear resistive network, where we assume the hard-margin SVM.

**Corollary 3.8.** *Let $u^* \in \mathbf{R}^m$ be the optimal solution of* [SVM]. *Then, for $i \in \{1, \dots, m\}$, $x_i$ is a support vector, i.e., $\eta_i u_i^* > 0$ if and only if there exists a path from $x_i$ to some $x_j$ with $\eta_i \neq \eta_j$ such that it contains no other labeled training vertex (data).*

Suppose that there exists some training data $x$ such that the deletion of $x$ from $(V, E)$ makes two or more connected components, i.e., $x$ is an articulation point of $(V, E)$. Let $U_1, \dots, U_k$ be the vertex sets of the connected components after the deletion of $x$. Let $(U_1 \cup \{x\}, E_1), \dots, (U_k \cup \{x\}, E_k)$ be subgraphs of $(V, E)$. Restricting training data set to each subgraph, we obtain SVM problems $[\mathrm{SVM}_1], \dots, [\mathrm{SVM}_k]$. From the electric network view point, this network can be regarded as a parallel connection of the component subgraphs. Hence the problem can naturally be decomposed. Translating this observation into a statement for the SVM problem, we see the following.

**Corollary 3.9.** *Under the above assumption, the optimal solution of* [SVM] *can be represented as the sum of optimal solutions of* $[\mathrm{SVM}_1], \dots, [\mathrm{SVM}_\mathrm{k}]$. *Consequently, for each $i \in \{1, \dots, k\}$, the restriction to $U_i \cup \{x\}$ of the decision function of the hard-margin* [SVM] *coincides with the decision function of* $[\mathrm{SVM}_i]$.

**Remark 3.10.** SVM with an electric network kernel falls in the scope of *discrete convex analysis* [15], which is a theory of convex functions with additional combinatorial structures. Specifically, the objective function of [SVM] with an electric network kernel is an *M-convex function* in continuous variables, and the optimization problem [SVM] is an M-convex function minimization problem.

**Remark 3.11.** Smola and Kondor [20] consider various kernels constructed from the Laplacian matrix $L$ of an undirected graph $(V, E)$. In particular, they introduced the kernel

$$K = (I + \sigma L)^{-1},$$

where $\sigma$ is a positive parameter. In our view, this kernel corresponds to the electric network kernel of a modified graph $(V \cup \{x_0\}, E \cup \{yx_0 \mid y \in V\})$ with a newly introduced root vertex $x_0$.

The computation of elements of $D$ or $K$ through numerical inversion of $\hat{L}$ is highly expensive because the size of $\hat{L}$ is usually very large. In Sections 4 and 5, we consider two classes of graphs $(V, E)$, trees and hypercubes, that admit efficient computation of the elements of $K$.

# 4 SVM on Trees

## 4.1 Relationship to Tree Metrics

In this section, we consider the case where $(V, E)$ is a tree. By regarding the resistance $r$ as the edge length, we then have

$$D(x, y) = \text{path length between } x \text{ and } y. \tag{4.1}$$

Hence $D$ is a tree metric. Recall [18] that a distance is called a *tree metric* if it can be expressed as the path length between vertices of some weighted tree. We take any $x_0 \in V$

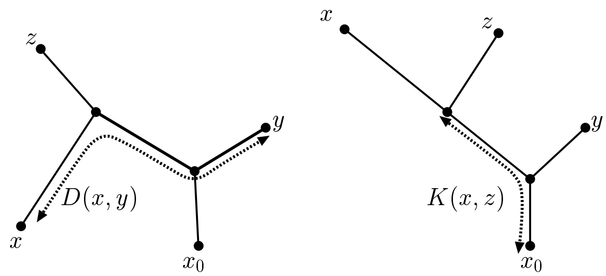


Figure 2: $D$ and $K$ on a tree

as the root. Then $K$ is given by

$$K(x,y) = \text{path length between } x_0 \text{ and the youngest common ancestor of } x \text{ and } y,$$

where "youngest" means "most distant from $x_0$." Hence $K$ can be recognized as the similarity function naturally derived from a *dendrogram* (see Figure 2).

As is well known, a tree metric can be characterized by the *four-point condition.*

**Theorem 4.1 ([2][19][23]).** *Let $\mathcal{X}$ be a finite set and $D : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ be a distance function on $\mathcal{X}$. Then $D$ is a tree metric if and only if $D$ satisfies the four-point condition:*

$$\begin{aligned}&\forall x, y, z, w \in \mathcal{X},\\ &D(x,y) + D(z,w) \leq \max\{D(x,z) + D(y,w), D(x,w) + D(y,z)\}.\end{aligned} \tag{4.2}$$

Corresponding to this, the following equivalent theorem is also well known; see [18] for example.

**Theorem 4.2.** *Let $\mathcal{X}$ be a finite set and $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ be a symmetric function on $\mathcal{X}$. Then $K$ is the Gromov product of some tree metric if and only if it satisfies the ultra-metric condition:*

$$\begin{aligned}&\forall x, y, z \in \mathcal{X},\\ &K(x,x) \geq K(x,y) \geq \min\{K(x,z), K(x,y)\} \geq 0.\end{aligned} \tag{4.3}$$

Hence, in any finite set $\mathcal{X}$, if we give a symmetric function $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}$ satisfying (4.3), then $\mathcal{X}$ is implicitly embedded into some weighted tree. In particular, $K$ is an electric network kernel. Hence the arguments in the previous section are applicable to SVM on $\mathcal{X}$ with this kernel $K$. Two specific applications of this idea are expounded below.

## 4.2 Min-Cut Kernel for Undirected Graphs

Let $G = (U, F, c)$ be an undirected graph with vertex set $U$, edge set $F$ and edge capacity $c : F \to \mathbf{R}_{>0}$. Let $\kappa : 2^U \to \mathbf{R}$ be the *cut function* of $G$ defined as

$$\kappa(X) = \sum\{c(e) \mid e = xy \in F, x \in X,\ y \in U \setminus X\} \quad (X \subseteq U). \tag{4.4}$$

We define the *min-cut kernel* $K : U \times U \to \mathbf{R}$ for $G$ as

$$K(x,y) = \begin{cases} \kappa(\{x\}) & \text{if } x = y \\ \min\{\kappa(X) \mid X \subseteq U, x \in X, y \in U \setminus X\} & \text{otherwise.} \end{cases}$$

**Proposition 4.3.** *The min-cut kernel $K$ satisfies the ultra-metric condition* (4.3).

*Proof.* Note that $\kappa$ satisfies $\kappa(X) = \kappa(U \setminus X) \geq 0$ for $X \subseteq V$. From this, nonnegativity of $K$ is observed. Clearly we have $K(x,x) = \kappa(\{x\}) \geq K(x,y)$ for $y \in U \setminus \{x\}$. We show $K(x,y) \geq \min\{K(x,z), K(y,z)\}$ for $x,y,z \in U$. Let $X^* \subseteq U$ be a minimizer of $\min\{\kappa(X) \mid X \subseteq U, x \in X, y \in U \setminus X\}$. Then we have $K(x,y) = \kappa(X^*)$. If $z \in X^*$, we have $\kappa(X^*) \geq K(y,z)$. If $z \notin X^*$, we have $\kappa(X^*) \geq K(x,z)$. □

Hence, the vertices of graph $G$ are implicitly embedded into some weighted tree by the min-cut kernel. The max-flow min-cut theorem implies that

$$K(x,y) = \text{maximum flow value between } x \text{ and } y \quad (x \neq y).$$

Hence, the value of $K$ can be efficiently computed through maximum flow algorithms or the Gomory-Hu cut tree algorithm [5].

**Remark 4.4.** The fact that the maximum flow value between two terminal node pair satisfies the ultra-metric condition is already known in 1960s [8], [12].

### $\boxed{4.3}$ Relationship to MPR Problem in Phylogeny

Here, we discuss the relationship between our SVM on trees and *Most-Parsimonious Reconstruction* (MPR) problem in phylogeny. First we briefly summarize the MPR problem; see [9], [14] for details. Let $\mathcal{C}$ be a set called the *character states*, and $d : \mathcal{C} \times \mathcal{C} \to \mathbf{R}$ be a distance function on $\mathcal{C}$. The MPR problem in phylogeny is mathematically formulated as follows:

> Given a tree $T = (V, E)$ (*phylogenetic tree*), a subset $X \subseteq V$, and a function $\chi : X \to \mathcal{C}$ called a *character* on $X$. Find a full character $\overline{\chi} : V \to \mathcal{C}$ that is a minimizer of the optimization problem
>
> $$\begin{aligned} [\text{MPR}] : \min_{\overline{\chi}: V \to \mathcal{C}} \quad & \sum \{ d(\overline{\chi}(x), \overline{\chi}(y)) \mid xy \in E,\ x, y \in V \} \\ \text{s.t.} \quad & \overline{\chi}(x) = \chi(x) \quad (x \in X). \end{aligned}$$

The following proposition indicates the relationship between our SVM on trees and the MPR problem.

**Proposition 4.5.** *Consider the hard-margin SVM on tree $T = (V, E)$ with unit resistance on each edge and training data set $(x_1, \eta_1), \ldots, (x_m, \eta_m) \subseteq V \times \{\pm 1\}$. Then the resulting decision function $f$ coincides with the solution of* [MPR] *problem with tree $T = (V, E)$, character state $\mathcal{C} = \mathbf{R}$, character $\chi : \{x_1, \ldots, x_m\} \to \mathbf{R}$ defined as $\chi(x_i) = \eta_i$, and distance function $d(u, v) = |u - v|^2$ for $u, v \in \mathbf{R}$.*

*Proof.* By Theorem 3.6 and Proposition 3.7, the hard-margin SVM problem is equivalent to the problem [FLOW] without inequality constraints. Hence its dual problem is given by

$$\begin{aligned} \min_{p: V \to \mathbf{R}} \quad & \frac{1}{2} \sum \{ (p(x) - p(y))^2 \mid xy \in E,\ x, y \in V \} \\ \text{s.t.} \quad & p(x_i) = \eta_i \quad (i = 1, \ldots, m). \end{aligned}$$

This problem coincides with [MPR] in the above. □

For various $\mathcal{C}$ and $d$, it is shown that MPR problems can be efficiently solved based on the dynamic programming [9], [14]. Hence our hard-margin SVM on trees can be also efficiently solved by these algorithms.

# 5 SVM on Hypercubes

## 5.1 Explicit Formula for the Resistance

In this section, we consider the case where $(V, E)$ is an $N$-dimensional hypercube. We regard $(V, E)$ as an electric network where all resistances of edges are equal to 1. Hence the node admittance matrix of $(V, E)$ coincides with the Laplacian matrix.

The vertices are naturally regarded as 0-1 vectors, i.e., $V = \{0,1\}^N$. Let $d_{\mathrm{H}} : V \times V \to \mathbf{R}$ be the Hamming distance defined as

$$d_{\mathrm{H}}(x,y) = \#\{i \in \{1,\dots,N\} \mid x^i \neq y^i\} \quad (x, y \in \{0,1\}^N),$$

or equivalently, $d_{\mathrm{H}}(x,y)$ is the minimum path length between $x$ and $y$ on $(V, E)$. It should be clear that $x^i$ denotes the $i$th element of $x$. By symmetry of the hypercube, the resistance $D$ between two vertex pair is given as a function in the Hamming distance of the pair as follows. The proof is presented in Subsection 5.3.

**Theorem 5.1.** *The resistance $D : V \times V \to \mathbf{R}$ of an $N$-dimensional hypercube $(V, E)$ is given by*

$$D(x,y) = \frac{1}{2^{N-2}} \sum_{s=1,3,5,\dots}^{d_{\mathrm{H}}(x,y)} \sum_{t=0}^{N-d_{\mathrm{H}}(x,y)} \frac{1}{2(s+t)} \binom{d_{\mathrm{H}}(x,y)}{s} \binom{N-d_{\mathrm{H}}(x,y)}{t}. \tag{5.1}$$

The theorem implies, in particular, that each element of kernel $K$ can be computed with $O(N^3)$ arithmetic operations. This makes it possible to apply the electric network kernel to large-scale practical problems on hypercubes.

**Remark 5.2.** The derivation of the explicit formula above relies essentially on the fact that the number of distinct eigenvalues of $L$ is bounded by $O(N)$; See Lemma 5.3 in Subsection 5.3. Similarly, the electric network kernel for $N$ tensor product of $k$-complete graph also admits an explicit formula, and hence can be efficiently computed because the Laplacian matrix for this graph has only $O(N)$ distinct eigenvalues [11].

## 5.2 Experimental Results

Here, we describe preliminary experiments with our electric network kernels on hypercubes. In order to estimate the performance, we compare the electric network kernel with the Hamming kernel and the diffusion kernel [13] using benchmark data having binary attributes. By the Hamming kernel we mean the kernel defined as

$$K(x,y) = N - d_{\mathrm{H}}(x,y) \quad (x, y \in \{0,1\}^N).$$

The diffusion kernel and the electric network kernel are implemented to LIBSVM package [3], which is one of the common SVM package programs. For benchmark data sets, we use `Hepatitis`, `Votes`, and `LED2-3` taken from UCI Machine Learning Repository [16] (Table 5.1). In `Hepatitis` data set, we use 12 binary attributes of all 20 attributes. `LED2-3` data set is made through the data generating tool in [16] by adding 10% noise.

Table 5.2 shows the experimental results with Hamming kernel (HK), diffusion kernel (DK), and electric network kernel (ENK) for these data sets, where Acc means the ratio of correct answers averaged over 40 random 5-fold cross validations and SVs is the number of support vectors for whole data set. Results are reported for the setting of the soft margin parameter $C$ and the diffusion coefficient $\beta$ achieving the best cross validated error rate.

Table 5.1: Data sets

| Data set | Size | Positive | Negative | Attribute |
|---|---|---|---|---|
| `Hepatitis` | 155 | 32 | 123 | 12 |
| `Votes` | 435 | 168 | 267 | 16 |
| `LED2-3` | 1914 | 937 | 977 | 7 |

Table 5.2: Experimental results

| | HK | | DK | | ENK | |
|---|---|---|---|---|---|---|
| Data set | SVs | Acc ($C$) | SVs | Acc ($C, \beta$) | SVs | Acc ($C$) |
| `Hepatitis` | 60 | 79.125 (64) | 60 | 79.775 (256, 3.5) | 106 | 77.725 (256) |
| `Votes` | 36 | 95.975 (4.0) | 53 | 96.025 (512, 3.0) | 274 | 84.450 (128) |
| `LED2-3` | 386 | 89.550 (2.0) | 392 | 89.700 (0.2, 3.0) | 388 | 89.800 (70) |

HK = Hamming kernel, DK = diffusion kernel, ENK = electric network kernel, SVs = number of support vectors, Acc = accuracy in percentile.

For `Hepatitis` and `Votes` data sets, three kernels show almost equivalent performance. For `Votes` data set, however, our ENK shows somewhat poor performance than others. In `Hepatitis` and `Votes`, ENK has larger SVs than other kernels. This phenomenon can be explained by Corollary 3.8 as follows. Since these two data sets are well separated than `LED2-3`, the soft margin SVM with ENK is close to the hard margin SVM. Hence it is expected from Corollary 3.8 that these SVM with ENK have many SVs.

The above results indicate that our electric network kernel works well as an SVM kernel. It is fair to say, however, that more extensive experiments against various kinds of data sets are required before its performance can be confirmed with more precision and confidence. Comprehensive computational study is left as a future research topic.

### 5.3 Proof of Theorem 5.1

First, we derive eigenvalues and eigenvectors of the Laplacian matrix $L_N$ of an $N$-dimensional hypercube. We regard functions defined on $\{0,1\}^N$ as $2^N$-dimensional vectors indexed by $\{0,1\}^N$ arranged in lexicographic order. Hence $L_N$ is an $2^N \times 2^N$ matrix like:

$$L_1 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad L_2 = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 2 & 0 & -1 \\ -1 & 0 & 2 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix}.$$

**Lemma 5.3.** *The eigenvalues of $L_N$ are given by*

$$2k \quad (k = 0, 1, \ldots, N) \tag{5.2}$$

*with the multiplicity of $2k$ being $\binom{N}{k}$. The eigenvectors for eigenvalue $2k$ are given by*

$$p^S = \phi_1^S \otimes \phi_2^S \otimes \cdots \otimes \phi_N^S \quad (S \subseteq \{1, 2, \ldots, N\},\ \#S = k), \tag{5.3}$$

*where $\phi_i^S$ is a 2-dimensional vector defined as*

$$\phi_i^S = \begin{cases} \begin{pmatrix} +1 \\ -1 \end{pmatrix} & \text{if } i \in S, \\ \begin{pmatrix} +1 \\ +1 \end{pmatrix} & \text{otherwise,} \end{cases} \qquad (i \in \{1, \ldots, N\}) \tag{5.4}$$

*and symbol $\otimes$ means Kronecker product.*

*Proof.* $L_N$ has the following recursive relation:

$$\begin{aligned} L_1 &= \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \\ L_{N+1} &= \begin{pmatrix} L_N + I & -I \\ -I & L_N + I \end{pmatrix}. \end{aligned} \tag{5.5}$$

From this, the characteristic polynomial $f_N(\lambda)$ of $L_N$ enjoys the following recursive relation:

$$\begin{aligned} & f_1(\lambda) = \lambda(\lambda - 2), \\ & f_{N+1}(\lambda) \\ & = \det(L_{N+1} - \lambda I) \\ & = \det \left[ \begin{pmatrix} I & 0 \\ -I & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & L_N - \lambda I \end{pmatrix} \begin{pmatrix} L_N + 2I - \lambda I & -I \\ 0 & I \end{pmatrix} \begin{pmatrix} I & 0 \\ I & I \end{pmatrix} \right] \\ & = \det(L_N - \lambda I) \det(L_N - (\lambda - 2) I) \\ & = f_N(\lambda) f_N(\lambda - 2). \end{aligned}$$

Hence, the characteristic polynomial $f_N(\lambda)$ of $L_N$ is given by

$$f_N(\lambda) = \prod_{k=0}^{N} (\lambda - 2k)^{\binom{N}{k}}. \tag{5.6}$$

Hence, we have (5.2). Next we consider eigenvectors. If $p$ is an eigenvector of eigenvalue $\lambda$ of $L_N$, then, from (5.5), we have

$$L_{N+1} \begin{pmatrix} p \\ p \end{pmatrix} = \lambda \begin{pmatrix} p \\ p \end{pmatrix}, \quad L_{N+1} \begin{pmatrix} p \\ -p \end{pmatrix} = (\lambda + 2) \begin{pmatrix} p \\ -p \end{pmatrix}. \tag{5.7}$$

The eigenvectors of $L_1$ are given as

$$\begin{pmatrix} +1 \\ +1 \end{pmatrix} \text{ for } \lambda = 0, \quad \begin{pmatrix} +1 \\ -1 \end{pmatrix} \text{ for } \lambda = 2. \tag{5.8}$$

From (5.7) and (5.8), we obtain (5.3). □

**Remark 5.4.** The graph of a hypercube can be expressed as a tensor product of single edges (1-dimensional hypercubes). Hence, Lemma 5.3 can be derived from the general formula for the spectra of the tensor product of graphs [6, Theorems 2.23 and 2.24 and p.75].

The expression (5.3) implies the following.

**Lemma 5.5.** *For $d \in \{1, \dots, N\}$, $1 \le i_1 < \cdots < i_s \le d$ and $d < j_1 < \cdots < j_t \le N$, we have*

$$p_{2^d}^{\{i_1, \dots, i_s, j_1, \dots, j_t\}} = (-1)^s, \tag{5.9}$$

*where the left-hand side above denotes the $2^d$th element of the vector $p^S$ with $S = \{i_1, \dots, i_s, j_1, \dots, j_t\}$.*

Let $P$ be a $2^N \times 2^N$ matrix whose column vectors are eigenvectors $p^S$, i.e.,

$$P = (p^S \mid S \subseteq \{1, 2, \dots, N\},\ \#S = k,\ 0 \le k \le N).$$

Then $L_N$ is diagonalized as

$$L_N = P \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_{2^N} \end{pmatrix} P^\top / 2^N,$$

where $0 = \lambda_1 < \lambda_2 \le \cdots \le \lambda_{2^N}$ are the eigenvalues of $L_N$. Let $\hat{L}_N$ and $\hat{P}$ be the $(2^N - 1) \times (2^N - 1)$ submatrices of $L_N$ and $P$, respectively, with the first columns and the first rows deleted. This means that $\mathbf{0} \in \{0, 1\}^N$ is taken as the root vertex $x_0$. Then we have

$$\hat{K} = (\hat{L}_N)^{-1} = 2^N (\hat{P})^{-1} \begin{pmatrix} 1/\lambda_2 & & \\ & \ddots & \\ & & 1/\lambda_{2^N} \end{pmatrix} (\hat{P}^\top)^{-1}. \tag{5.10}$$

**Lemma 5.6.** *$\hat{P}^{-1}$ is given as*

$$\hat{P}^{-1} = (\hat{P}^\top - \mathbf{1}\mathbf{1}^\top)/2^N, \tag{5.11}$$

*where $\mathbf{1}$ is the $(2^N - 1)$-dimensional vector with all elements 1.*

*Proof.* $P$ can be expressed as

$$P = \begin{pmatrix} 1 & \mathbf{1}^\top \\ \mathbf{1} & \hat{P} \end{pmatrix}.$$

Since $PP^\top = 2^N I$, we have

$$\mathbf{1}^\top + \mathbf{1}^\top \hat{P}^\top = 0, \quad \mathbf{1}\mathbf{1}^\top + \hat{P}\hat{P}^\top = 2^N I. \tag{5.12}$$

Substituting the first equation to the second in (5.12), we obtain

$$-\mathbf{1}\mathbf{1}^\top \hat{P}^\top + \hat{P}\hat{P}^\top = 2^N I.$$

This implies $\hat{P}^{-1} = (\hat{P}^\top - \mathbf{1}\mathbf{1}^\top)/2^N$. □

It follows from (5.10) and (5.11) that

$$\hat{K} = (\hat{P} - \mathbf{1}\mathbf{1}^\top) \begin{pmatrix} 1/\lambda_2 & & \\ & \ddots & \\ & & 1/\lambda_{2^N} \end{pmatrix} (\hat{P} - \mathbf{1}\mathbf{1}^\top)^\top / 2^N. \tag{5.13}$$

Finally we derive the resistance $D$. For $x, y \in \{0,1\}^N$ with $d = d_{\mathrm{H}}(x,y)$, from symmetry of the hypercube, we have

$$D(x,y) = D(0, 2^d) = \hat{K}(2^d, 2^d),$$

where $2^d$ is a short-hand notation for $(\underbrace{1,\cdots,1}_{d},\underbrace{0,\cdots,0}_{N-d})$. From (5.13) and (5.9), we have

$$\begin{aligned}
\hat{K}(2^d, 2^d) &= \frac{1}{2^N}\sum_{i=2}^{2^N}\frac{1}{\lambda_i}(P_{i2^d}-1)^2 \\
&= \frac{1}{2^N}\sum_{k=1}^{N}\frac{1}{2k}\sum_{1\le i_1<i_2<\cdots<i_k\le N}(p_{2^d}^{\{i_1,i_2,\ldots,i_k\}}-1)^2 \\
&= \frac{1}{2^N}\sum_{k=1}^{N}\sum_{\substack{0\le s,\ 0\le t\\ s+t=k}}\frac{1}{2(s+t)}\sum_{\substack{1\le i_1<\cdots<i_s\le d\\ d<j_1<\cdots<j_t\le N}}((-1)^s-1)^2 \\
&= \frac{1}{2^{N-2}}\sum_{s=1,3,5,\ldots}^{d}\sum_{t=0}^{N-d}\frac{1}{2(s+t)}\binom{d}{s}\binom{N-d}{t}.
\end{aligned}$$

Thus we have proven Theorem 5.1.

## Acknowledgements

The authors thank Satoru Iwata and Shiro Matuura for helpful suggestions concerning the formula in Theorem 5.1, and anonymous referees for suggestive comments. This work is an outcome of a joint project between University of Tokyo and Justsystem Co. This work is also supported by the 21st Century COE Program on Information Science and Technology Strategic Core, and a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

## References

[1] A. Berman and R.J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, SIAM, Philadelphia, 1994.

[2] P. Buneman, The recovery of trees from measures of dissimilarity, in *Mathematics in the Archaeological and Historical Sciences*, F.R. Hodson, D.G. Kendall and P. Tautu (eds.), Edinburgh University Press, Edinburgh, 1971, pp. 387–395.

[3] C.C. Chang and C.J. Lin, LIBSVM a library for support vector machines, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[4] F. Chung and S.-T. Yau, Discrete Green's functions, *Journal of Combinatorical Theory Series A* 91 (2000) 191–214.

[5] W.J. Cook, W.H. Cunningham, W.R. Pulleyblank and A. Schrijver, *Combinatorial Optimization*, Wiley, New York, 1998.

[6] D.M. Cvetković, M. Doob and H. Sachs, *Spectra of Graphs*, 3rd. ed., Johann Ambrosius Barth Verlag, Heidelberg–Leipzig, 1995.

[7] M. Fiedler, Some characterizations of symmetric inverse $M$-matrices, *Linear Algebra and Its Applications* 275/276 (1998) 179–187.

[8] L.R. Ford, Jr., and D.R. Fulkerson, *Flows in Networks*, Princeton University Press, Princeton, N.J., 1962.

[9] M. Hanazawa, H. Narushima and N. Minaka, Generating most parsimonious reconstructions on a tree: a generalization of the Farris-Swofford-Maddison method, *Discrete Applied Mathematics* 56 (1995) 245–265.

[10] D. Haussler, Convolution kernels on discrete structures, Technical report UCSC-CRL-99-10, Department of Computer Science, University of California at Santa Cruz, 1999.

[11] H. Hirai, K. Murota and M. Rikitoku, Electric network kernel for support vector machines, Decision Theory and Optimization Algorithms, *RIMS Kokyuroku* 1409 (2005) 205–218.

[12] M. Iri, *Network Flow, Transportation and Scheduling*, Academic Press, New York, 1969.

[13] R.I. Kondor and J. Lafferty, Diffusion kernels on graphs and other discrete structures, in *Proceedings of the 19th International Conference on Machine Learning*, B. Schölkopf and M. Warmuth (eds.), Morgan Kaufmann, San Fransisco, CA, 2002, pp. 315–322.

[14] N. Minaka, *Systematics, Phylogenetics, and the Tree of Life: A Cladistic Perspective*, University of Tokyo Press, Tokyo, 1997 (In Japanese).

[15] K. Murota, *Discrete Convex Analysis*, SIAM, Philadelphia, PA, 2003.

[16] P.M. Murphy and D.W. Aha, *UCI Repository of machine learning databases*, `http://www.ics.uci.edu/~mlearn/MLRepository.html`, Irvine, CA: University of California, Department of Information and Computer Science, 1994.

[17] B. Schölkopf and A.J. Smola, *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.

[18] C. Semple and M. Steel, *Phylogenetics*, Oxford University Press, Oxford, 2003.

[19] J.M.S. Simões-Pereira, A note on the tree realizability of a distance matrix, *Journal of Combinatorial Theory* 6 (1969) 303–310.

[20] A.J. Smola and R. Kondor, Kernels and regularization on graphs, in *Proceedings of the Annual Conference on Computational Learning Theory and Kernel Workshop*, B. Schölkopf and M. Warmuth (eds.), Lecture Notes in Computer Science 2777, Springer, 2003.

[21] V. Vapnik, *Statistical Leaning Theory,* Wiley, 1998.

[22] C. Watkins, Dynamic alignment kernels, in *Advances in Large-Margin Classifiers,* A.J. Smola, B. Schölkopf, P. Barlett and D. Schuurmans (eds.), MIT Press, Cambridge, MA, 2000, pp. 39–50.

[23] K.A. Zareckiĭ, Constructing a tree on the basis of a set of distances between the hanging vertices, *Uspehi Matematičeskih Nauk* 20 (1965) 90–92.

*Manuscript received 31 August 2004*
*revised 8 February 2005*
*accepted for publication 12 February 2005*

HIROSHI HIRAI
Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606-8502, Japan
E-mail address: hirai@kurims.kyoto-u.ac.jp

KAZUO MUROTA
Department of Mathematical Informatics,
Graduate School of Information Science and Technology, University of Tokyo, Tokyo 113-8656, Japan
E-mail address: murota@mist.i.u-tokyo.ac.jp

MASAKI RIKITOKU
Justsystem Corporation Innovative Technology R&D Strategy Dept., Aoyama bldg. 1-2-3, Kita-Aoyama, Tokyo 107-8640, Japan
E-mail address: Masaki_Rikitoku@justsystem.co.jp