An ILP Formulation for Integral Cycle Basis
personal communication by Andreas Karrenbauer

Let $G = (V,E)$ be a connected undirected graph and let $w : E \to \mathbb{R}_{\geq 0}$ be a non-negative weight function. Recall that a basis is integral if any circuit is an integral linear combination of the circuits in the basis. Let T be an arbitrary spanning tree of G and let N be the co-tree arcs with respect to T. For a cycle basis B, let Γ_B be its cycle matrix and let Γ'_B be the square submatrix selected by the co-tree arcs.

Lemma 1 *B is an integral basis iff* $\det \Gamma'_B = \pm 1$.

Proof: If B is an integral basis then any fundamental circuit with respect to T is an integral linear combination of the circuits in B, i.e., there is an integral matrix A such that $\Phi = \Gamma_B A$, where Φ is the cycle matrix of the fundamental basis. Restricting to the rows in N yields $I = \Gamma'_B A$ and hence $1 = \det \Gamma'_B \det A$. Since A is integral, we have $\det \Gamma'_B = \pm 1$ and $\det A = \pm 1$.

Conversely, assume $\det \Gamma'_B = \pm 1$ and let C be any circuit. Then $C = \Gamma_B x_C$; x_C is integral by Cramer's rule. ∎

The Lemma implies that deciding whether there is an integral basis of weight at most K is in NP. We guess a basis, verify that its determinant is one and that is weight is at most K. Integer linear programming is NP-complete. Any problem in NP can be reduced to any NP-complete problem. Therefore there must be a formulation of the minimum integral cycle basis problem as an ILP. Andreas Karrenbauer suggested the following formulation.

Let $\nu = m - n + 1$. We have integral variables $C_{i,e}$ for $1 \leq i \leq \nu$ and $e \in E$. $C_{i,e}$ indicates the usage of edge e in circuit C_i. Since, the C_i are circuits, $-1 \leq C_{i,e} \leq 1$.

The objective is to minimize $\sum_{i,e} w(e)|C_{i,e}|$.

We first argue how to deal with the absolute values in the objective function. The inequalities $C_{i,e} \leq A_{i,e}$ and $-C_{i,e} \leq A_{i,e}$ guarantee $|C_{i,e}| \leq A_{i,e}$. Let us try the objective $\sum_{i,e} w(e) A_{i,e}$. Since we are minimizing, we are guaranteed $A_{i,e} = |C_{i,e}|$ for the edges e with $w(e) > 0$. For the edges with $w(e) = 0$, we do not have this guarantee. We can enforce it by choosing a large number M and changing the objective to $M \sum_{i,e} w(e) A_{i,e} + \sum_{i,e} A_{i,e}$. Assume that the $w(e)$ are integers. Then distinct object values differ by at least 1; after multiplying by M, they differ by at least M. Therefore, if M is larger than m^2, the term $\sum_{i,e} A_{i,e}$ cannot turn a non-optimal solution into an optimal solution. If the $w(e)$ are reals, we cannot choose a concrete value for M. We treat it as a symbolic value (an infimaximal).

We come to the constraints. The first requirement is that the C_i's are cycles. We have for every i and v, the flow conservation constraint

$$\sum_{e \in \delta^+(v)} C_{i,e} = \sum_{e \in \delta^-(v)} C_{i,e} \ .$$

The second requirement is that the C_i form an integral basis. We formulate that the cycle matrix has an integral inverse. Let N be the co-tree arcs. We have variables $D_{i,e}$ for $1 \leq i \leq \nu$ and $e \in N$ and the constraints

$$\sum_{e \in N} C_{i,e} D_{j,e} = \delta_{ij} \ ,$$

where δ_{ij} is zero for $i \neq j$ and 1 otherwise. However, this constraint is not a linear constraint. We will next show how to express this constraint by a set of linear constraints. We introduce the following intermediate values.

1. $C_{i,e}D_{j,e} = F_{i,j,e}$.
2. $\sum_e F_{i,j,e} = \delta_{ij}$.

The second condition is linear. For the first condition, we need to go to the digit level. Assume we know an upper bound K on the absolute values of the entries of matrix D (see below) and let $K < 2^k$. We represent integers Z with $|Z| < K$ in 1-complement, i.e.,

$$Z = \sum_{0 \leq \ell < k} Z_\ell 2^\ell - Z_k(2^k - 1)$$

with $Z_\ell \in \{0,1\}$ for $0 \leq \ell \leq k$. Observe that the constraints that relate Z to its digits are linear constraints.

We can now formulate the first constraint. Define $X_{i,e}$ and $Y_{i,e}$ by the constraints

$$2X_{i,e} = C_{i,e} + A_{i,e} \text{ and } 2Y_{i,e} = -C_{i,e} + A_{i,e} .$$

Then $X_{i,e}, Y_{i,e} \in \{0,1\}$, $X_{i,e} = (C_{i,e} = 1)$ and $Y_{i,e} = (C_{i,e} = -1)$. Therefore $F_{i,j,e} = X_{i,e}D_{j,e} - Y_{i,e}D_{j,e}$.

We next observe that multiplication of single binary digits can be written as linear conditions. For $a, b, c \in \{0,1\}$, we have $a \cdot b = c$ iff $c \leq a$, $c \leq b$ and $c \geq a + b - 1$. Let $D^\ell_{j,e}$, $XD^\ell_{i,j,e}$, and $YD_{i,j,e}$, $0 \leq \ell \leq k$ be the digits of $D_{j,e}$, $X_{i,e}D_{j,e}$, and $Y_{i,e}D_{j,e}$, respectively. Then

$$XD^\ell_{i,j,e} \leq D^\ell_{j,e} \quad XD^\ell_{i,j,e} \leq X^\ell_{i,e} \quad XD^\ell_{i,j,e} \geq D^\ell_{j,e} + X^\ell_{i,e} - 1 \quad XD_{i,j,e} = \sum_{0 \leq \ell < k} XD^\ell_{i,j,e} 2^\ell - XD^k_{i,j,e}(2^k - 1)$$

and similarly for the YD's.

Lemma 2 *The entries of D are bounded by $m!$*

Proof: D is the inverse of a matrix with entries in $\{0,1,-1\}$. Any entry of the inverse is (up to sign) the determinant of a minor. The determinant of a minor is at most $m!$ (since the determinant of a $m \times m$ matrix is the sum of $m!$ terms and each term is the product of matrix entries). ∎

Theorem 1 *The above is a ILP formulation of the minimum integral cycle basis problem.*

Since $m! \leq m^m$, we can work with $k = m \log m$.

Andreas told me he can compute the optimal integral basis of the Petersen graph $P_{7,2}$ (7 nodes, 21 edges) in about 5 minutes. For the details, see his email below.

I translate part of it into German. Andreas observes that there is no need to introduce an M into the objective function. First, for edges with $w(e) = 0$, the value of $A(i,e)$ is irrelevant. Second, he gives an alternative definition of the variables $X_{i,e}$ and $Y_{i,e}$, namely

$$C_{i,e} = X_{i,e} - Y_{i,e} \quad \text{and} \quad X_{i,e} + Y_{i,e} \leq 1 .$$

Lieber Kurt,

ich habe mir Deinen Aufschrieb durchgelesen. Gibt es mittlerweile neue Erkenntnisse, dass man sich auf circuits fuer die minimale ganzzahlige Kreisbasis beschraenken kann? Der ILP-Ansatz funktioniert abgewandelt auch fuer
den allgemeinen Fall.

Wenn ich das jetzt richtig verstanden habe, dann fuehrst Du das M in die Zielfunktion ein, damit A_{i,e} immer der Betrag von C_{i,e} ist, um damit X_{i,e} und Y_{i,e} zu definieren. In meinen Augen ist das nicht unbedingt noetig, wenn man C_{i,e} = X_{i,e} - Y_{i,e} als constraints nimmt, sowie X_{i,e} + Y_{i,e} \le 1.

Mit einer Aussage ueber die Groesse der loesbaren Instanzen moechte ich vorsichtig sein. Zum Einen haengt die Anzahl der binaeren Variablen von m - n + 1 ab (quadratisch) und von dem benutzten k (linear). Zum Anderen kommt es auch stark auf die Struktur des Graphen an. Ich habe bis jetzt auch noch keine ausgefallene Heuristik fuer die Startloesung implementiert (es wird einfach eine beliebige fundamentale Basis generiert und dann mit der Einheitsmatrix gestartet). Zum Beispiel kann ich die optimale Basis fuer der verallgemeinerten Petersen-Graph P_{7,2} (21 Kanten, 7 Knoten) in ca. 5 min berechnen, wenn K=2 gesetzt wird; das 4x4 grid in ca. 1min, obwohl die Startloesung dort viel weiter vom Optimum entfernt liegt.

Ich habe Dir das Demo-Programm mit GUI angehaengt. Es laesst sich z.B. auf "dork" mit

g++ -g -O3 -fPIC -fexceptions -DIL_STD -I /LEDA/LEDA-6.x/LEDA-6.2.1-std/linux/g++-4.1/incl/ icbguibinary.cc -o icbguibinary -L /LEDA/LEDA-6.x/LEDA-6.2.1-std/linux/g++-4.1/ -lleda -lilocplex -lcplex -lconcert -lpthread -lm -lX11

kompilieren und dann mit ./icbguibinary starten. Das B gibt die Anzahl der bits
in 2-complement an. Klicken auf "done" startet die Berechnung fuer den aktuellen Graphen. Alle Informationen werden nach std out geschrieben.

Eine interessante Frage waere, wieviele Bits fuer die Inverse ausreichen (konstant?).

Beste Gruesse,

Andreas