RUTCOR RESEARCH REPORTS, 2013

RUTCOR – Rutgers Center for Operations Research
Busch Campus, Rutgers University
640 Bartholomew Road, Piscataway, NJ 08854-8003

RRR 1-2013 **ON CIS CIRCULANTS**, Endre Boros, Vladimir Gurvich, Martin Milani

A *circulant* is a Cayley graph over a cyclic group. A *well-covered* graph is a graph in which all maximal stable sets are of the same size $\alpha = \alpha(G)$, or in other words, they are all maximum. A *CIS* graph is a graph in which every maximal stable set and every maximal clique intersect. It is not difficult to show that a circulant G is a CIS graph if and only if G and its complement $\overline{G}$ are both well-covered and the product $\alpha(G)\alpha(\overline{G})$ is equal to the number of vertices. It is also easy to demonstrate that both families, the circulants and the CIS graphs, are closed with respect to the operations of taking the complement and lexicographic product. We study the structure of the CIS circulants. It is well-known that all P_4-free graphs are CIS. In this paper, in addition to the simple family of the P_4-free circulants, we construct a non-trivial sparse but infinite family of CIS circulants. We are not aware of any CIS circulant that could not be obtained from graphs in this family by the operations of taking the complement and lexicographic product.

Keywords: Circulant, CIS graph, well-covered graph, maximal stable set; maximum stable set; maximal clique; maximum clique.
Math. Subj. Class. (2010): 05C25, 05C69

RRR 2-2013 **LARGE MARGIN CASE-BASED REASONING**, Martin Anthony, Joel Ratsaby

The central problem in case based reasoning (CBR) is to infer a solution for a new problem-instance by using a collection of existing problem-solution cases. The basic heuristic guiding CBR is the hypothesis that similar problems have similar solutions. CBR has been often criticized for lacking a sound theoretical basis, and there has only recently been some attempts at formalizing CBR in a theoretical framework, including work by Hullermeier who made the link between CBR and the probably approximately correct (PAC) theoretical model of learning in his 'case-based inference' (CBI) formulation. In this paper we present a new framework of CBI which models it as a multi-category classification problem. We use a recently-developed notion of geometric margin of classification to obtain generalization error bounds.

RRR 3-2013 **GRAPHS WHOSE COMPLEMENT AND SQUARE ARE ISOMORPHIC (EXTENDED VERSION)**, Martin Milanič, Anders Sune Pedersen, Daniel Pellicer, Gabriel Verret

We study square-complementary graphs, that is, graphs whose complement and square are isomorphic. We prove several necessary conditions for a graph to be square-complementary, describe ways of building new square-complementary graphs from existing ones,construct infinite families of square-complementary graphs, and characterize square-complementary graphs within various graph classes. The bipartite case turns out to be of particular interest.

RRR 4-2013 **BOTTOM-UP APPROACH TO MEASURE RISK FOR DECISION-MAKING ON CORPORATE MERGERS AND ACQUISITIONS (M&AS)**, Jinwook Lee, András Prékopa

Corporate Mergers and Acquisitions (M&As) are notoriously complex, and risk management is one of the essential aspects of the analysis process for the decision-making on M&A deals. Empirically, we see that some M&A transactions are not successful in part because of the increased exposure to correlated sectors, suggesting that the merged entity possesses increased risk (strategic, operational, financial, etc.) in the market. This motivates our research on risk evaluation processes for corporate M&A deals, and in particular for a suitable, but not necessarily convex, risk measure for multiple correlated assets capturing their stochastic dependence structure. In this paper, such multivariate risk measure is introduced and surveyed. Numerical examples are presented.

Keywords: Corporate Mergers and Acquisitions; Risk Management; Multivariate Risk; Decision Analysis

RRR 5-2013 **LOGICAL ANALYSIS OF MULTI-CLASS DATA**, Juan Felix Avila Herrera, Munevver Mine Subasi

Logical Analysis of Data (LAD) is a two-class learning algorithm which integrates principles of combinatorics, optimization, and the theory of Boolean functions. This paper proposes an algorithm based on mixed integer linear programming to extend the LAD methodology to solve multi-class classification problems, where One-vs-Rest (OvR) learning models are efficiently constructed to classify observations in predefined classes. The utility of the proposed approach is demonstrated through experiments on multi-class benchmark datasets.

RRR 6-2013 **SHARP BOUNDS OF THE MULTIVARIATE DISCRETE MOMENT PROBLEM: APPROXIMATIONS OF CDF VALUES AND RELIABILITY LEVELS**, Gergely Mádi-Nagy, Zoltán Csaba Nagy

The method of polynomial bases in Mádi-Nagy (2012) can solve almost any type of multivariate discrete moment problem (MDMP) within an acceptable time. The effective implementation of this algorithm is published as a Google hosted project Numerical MDMP (2013), which enables the use of the method in several practical applications. This paper presents three types of applications: approximations of cumulative distribution function (CDF) values of discrete and continuous random vectors and bounding network reliabilities. In case of CDF of discrete random vectors a new bounding method is presented with useful results. For the CDF of continuous random vector our method the most suitable for two-sided bounding among the investigated ones. As regards the network reliability, the new bounds are much better than the other bounds of the comparison. Hence, the polynomial algorithm with its implementation seems to be a useful tool for probability bounding.

Keywords: Multivariate discrete moment problem, Linear programming, Expectation bounds, Probability bounds, Multivariate distribution function, Network reliability

AMS: 30E05, 90C05, 60E15, 60K10, 62H99

RRR 7-2013 **TIME CONSISTENCY VERSUS LAW INVARIANCE IN MULTISTAGE STOCHASTIC OPTIMIZATION WITH COHERENT RISK MEASURES: MULTILEVEL OPTIMIZATION MODELING AND COMPUTATIONAL COMPLEXITY**, Jonathan Eckstein

Coherent risk measures have become a popular tool for incorporating risk aversion into stochastic optimization models. For dynamic models in which uncertainly is resolved at more than one stage, however, use of coherent risk measures within a standard single-level optimization framework presents the modeler with an uncomfortable choice between two desirable model properties, time consistency and law invariance. Prior published work has favored maintaining time consistency, but the absence of law invariance makes the resulting models unattractive to practical decision makers. This paper summarizes these issues and then presents an alternative multilevel optimization modeling approach that preserves law invariance, yet leads to *models* that are time-consistent even while using time-inconsistent risk measures. It argues that this approach should be the starting point for all multistage optimization modeling; however, when performing classical risk-neutral modeling, it simplifies to a more familiar single-objective form.

Unfortunately, this paper also shows that its proposed approach leads to $\mathcal{NP}$-hard models, even in the simplest imaginable setting in which it would be needed: three-stage linear problems on a finite probability space, using the standard mean-semideviation and average-value-at-risk measures. While not necessarily indicating that solution of such models is impractical, these results suggest that it will likely require approximation or implicit enumeration methods. We close with some preliminary computational results showing that high-quality local optimum solutions of models of the kind we propose are in fact practically computable, hence that the complexity results should not be taken as completely discouraging.

RRR 8-2013 **PRICE-BANDS: A TECHNICAL TOOL FOR STOCK TRADING**, Jinwook Lee, Joonhee Lee, András Prékopa

Given a stochastic process with known finite dimensional distributions, we construct lower and upper bounds within which future values of the stochastic process run, at a fixed probability level. For a financial trading business, such set of bounds are called price-bands or trading-bands that can be used as an indicator for successfully buying or short-selling shares of stock. In this paper, we present a mathematical model for the novel construction of price-bands using a stochastic programming formulation. Numerical examples using recent US stock market intraday data are presented.

Keywords. Price-bands (trading-bands); Technical Analysis; Gaussian Process; Binomial Moment Problem

RRR 9-2013 **PEBBL: AN OBJECT-ORIENTED FRAMEWORK FOR SCALABLE PARALLEL BRANCH AND BOUND**, Jonathan Eckstein, William E. Hart, Cythia A. Phillips

PEBBL is a C++ class library implementing the underlying operations needed to support a wide variety of branch-and-bound algorithms in a message-passing parallel computing environment. Deriving application-specific classes from PEBBL, one may create parallel branch-and-bound applications through a process focused on the unique aspects of the application, while relying on PEBBL for generic aspects of branch and bound, such as

managing the active subproblem pool across multiple processors, load balancing, and termination detection. PEBBL is designed to provide highly scalable performance on large numbers of processor cores, using a distributed-memory programming model and MPI message passing.

We describe the basics of PEBBL's architecture and usage, with emphasis on the library's key innovations and contributions, including the notion of branch and bound as a set of operators that manipulate subproblem state transitions. We also describe PEBBL's special features, such as parallel checkpointing, support for specialized ramp-up procedures, and the ability to exhaustively enumerate specified sets of near-optimal solutions.

Finally, we present an example application, the maximum monomial agreement (MMA) problem arising in machine learning applications. For sufficiently difficult problem instances, we show essentially linear speedup on over 6,000 processor cores. We also show how processor cache effects can produce reproducibly superlinear speedups.

RRR 10-2013 **GENERATING MINIMAL VALID INEQUALITIES**, Emre Yamangil, Endre Boros

We consider the problem of generating a lattice-free convex set to find a valid inequality that minimizes the sum of its coefficients for 2-row simplex cuts. Multi-row simplex cuts has been receiving considerable attention recently and we show that a pseudo-polytime generation of a lattice-free convex set is possible. We conclude with a short numerical study.

RRR 11-2013 **GENERALIZED SANDWICH PROBLEM FOR Π- AND Δ-FREE MULTIGRAPHS AND ITS APPLICATIONS TO POSITIONAL EXTENSIONS OF GAME FORMS**, Endre Boros, Vladimir Gurvich

An *n-multigraph* $\mathcal{G} = (V; E_i \mid i \in I)$ is a complete graph $G = (V, E)$ whose edges are covered by $n = |I|$ sets $E = \cup_{i \in I} E_i$, some of which might be empty. If this cover is a partition, then we call $\mathcal{G}$ an *n-graph*. We say that an n-graph $\mathcal{G}' = (V; E'_i \mid i \in I)$ is an edge-subgraph of an n-multigraph $\mathcal{G} = (V; E_i \mid i \in I)$ if $E'_i \subseteq E_i$ for all $i \in I$. We denote by Δ the n-graph on three vertices with three nonempty sets each containing a single edge, and by Π the four-vertex n-graph with two non-empty sets each of which contains the edges of a P_4. In this paper, we recognize in polynomial time whether a given n-multigraph $\mathcal{G}$ contains a Π- and Δ-free n-subgraph, or not, and if yes provide a polynomial delay algorithm to generate all such subgraphs. The above decision problem can be viewed as a direct generalization of the sandwich problem for P_4-free graphs introduced and solved by Golumbic, Kaplan, and Shamir in 1995.

As a motivation and application, we consider the n-person positional game forms, which are known to be in a one-to-one correspondence with Π- and Δ-free n-graphs. Given game form g, making use of the above result, we recognize in polynomial time whether g is a subform of a positional (that is, tight and rectangular) game form and, if yes, we generate with polynomial delay all such positional extensions of g.

Keywords: edge-colored n-graph, Gallai graph, sandwich problem, positional, tight, and rectangular game forms; polynomial delay.

RRR 12-2013 **QUADRATIZATION OF SYMMETRIC PSEUDO-BOOLEAN FUNCTIONS**, Martin Anthony, Endre Boros, Yves Crama, Aritanan Gruber

A *pseudo-Boolean function* is a real-valued function $f(x) = f(x_1, x_2, \ldots, x_n)$ of n binary variables; that is, a mapping from $\{0,1\}^n$ to $\mathbf{R}$. For a pseudo-Boolean function $f(x)$ on $\{0,1\}^n$, we say that $g(x,y)$ is a *quadratization* of f if $g(x,y)$ is a quadratic polynomial depending on x and on m *auxiliary* binary variables $y_1, y_2, \ldots, y_m$ such that $f(x) = \min\{g(x,y) : y \in \{0,1\}^m\}$ for all $x \in \{0,1\}^n$. By means of quadratizations, minimization of f is reduced to minimization (over its extended set of variables) of the quadratic function $g(x,y)$. This is of some practical interest because minimization of quadratic functions has been thoroughly studied for the last few decades, and much progress has been made in solving such problems exactly or heuristically. A related paper [**?**] initiated a systematic study of the minimum number of auxiliary y-variables required in a quadratization of an arbitrary function f (a natural question, since the complexity of minimizing the quadratic function $g(x,y)$ depends, among other factors, on the number of binary variables). In this paper, we determine more precisely the number of auxiliary variables required by quadratizations of *symmetric* pseudo-Boolean functions $f(x)$, those functions whose value depends only on the Hamming weight of the input x (the number of variables equal to 1).