The Design of Competitive Online Algorithms via a Primal–Dual Approach

The Design of Competitive Online Algorithms via a Primal–Dual Approach

Niv Buchbinder

Technion — Israel Institute of Technology
Israel
nivb@cs.technion.ac.il

Joseph (Seffi) Naor

Technion — Israel Institute of Technology
Israel
naor@cs.technion.ac.il

the essence of knowledge

Boston – Delft

Foundations and Trends® in Theoretical Computer Science

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is N. Buchbinder and J. (Seffi) Naor, The Design of Competitive Online Algorithms via a Primal–Dual Approach, Foundation and Trends® in Theoretical Computer Science, vol 3, nos 2–3, pp 93–263, 2007

ISBN: 978-1-60198-216-2

Foundations and Trends® in Theoretical Computer Science
Volume 3 Issue 2–3, 2007

Editorial Board

Editorial Scope

Foundations and Trends® in Theoretical Computer Science will publish survey and tutorial articles in the following topics:

- Algorithmic game theory
- Computational algebra
- Computational aspects of combinatorics and graph theory
- Computational aspects of communication
- Computational biology
- Computational complexity
- Computational geometry
- Computational learning
- Computational Models and Complexity
- Computational Number Theory
- Cryptography and information security
- Data structures
- Database theory
- Design and analysis of algorithms
- Distributed computing
- Information retrieval
- Operations Research
- Parallel algorithms
- Quantum Computation
- Randomness in Computation

Information for Librarians

Foundations and Trends® in Theoretical Computer Science, 2007, Volume 3, 4 issues. ISSN paper version 1551-305X. ISSN online version 1551-3068. Also available as a combined paper and online subscription.

Foundations and Trends® in
Theoretical Computer Science
Vol. 3, Nos. 2–3 (2007) 93–263

DOI: 10.1561/0400000024

The Design of Competitive Online Algorithms via a Primal–Dual Approach

Niv Buchbinder[1] and Joseph (Seffi) Naor[2]

[1] *Computer Science Department, Technion — Israel Institute of Technology, Israel, nivb@cs.technion.ac.il*
[2] *Computer Science Department, Technion — Israel Institute of Technology, Israel, naor@cs.technion.ac.il*

Abstract

The primal–dual method is a powerful algorithmic technique that has proved to be extremely useful for a wide variety of problems in the area of approximation algorithms for NP-hard problems. The method has its origins in the realm of exact algorithms, e.g., for matching and network flow. In the area of approximation algorithms, the primal–dual method has emerged as an important unifying design methodology, starting from the seminal work of Goemans and Williamson [60].

We show in this survey how to extend the primal–dual method to the setting of online algorithms, and show its applicability to a wide variety of fundamental problems. Among the online problems that we consider here are the weighted caching problem, generalized caching, the set-cover problem, several graph optimization problems, routing, load balancing, and the problem of allocating ad-auctions. We also show that classic online problems such as the ski rental problem and the dynamic TCP-acknowledgement problem can be solved optimally using a simple primal–dual approach.

The primal–dual method has several advantages over existing methods. First, it provides a general recipe for the design and analysis of online algorithms. The linear programming formulation helps detecting the difficulties of the online problem, and the analysis of the competitive ratio is direct, without a potential function appearing "out of nowhere." Finally, since the analysis is done via duality, the competitiveness of the online algorithm is with respect to an optimal fractional solution, which can be advantageous in certain scenarios.

Contents

1

Introduction

The primal–dual method is a powerful algorithmic technique that has proved to be extremely useful for a wide variety of problems in the area of approximation algorithms. The method has its origins in the realm of exact algorithms, e.g., for matching and network flow. In the area of approximation algorithms, the primal–dual method has emerged as an important unifying design methodology starting from the seminal work of Goemans and Williamson [60].

Our goal in this survey is to extend the primal–dual method to the setting of online algorithms, and show that it is applicable to a wide variety of problems. The approach we propose has several advantages over existing methods:

- A general recipe for the design and analysis of online algorithms is developed.
- The framework is shown to be applicable to a wide range of fundamental online problems.
- A linear programming formulation helps detecting the difficulties of the online problem in hand.
- The competitive ratio analysis is direct, without a potential function appearing "out of nowhere."

- The competitiveness of the online algorithm is with respect to an optimal fractional solution.

In Section 2, we briefly provide the necessary background needed for the rest of the discussion. This includes a short exposition on linear programming, duality, offline approximation methods, and basic definitions of online computation. Many readers may already be familiar with these basic definitions and techniques; however, we advise the readers not to skip this chapter, and in particular the part on approximation algorithms. Techniques pertinent to approximation algorithms are presented in a way that allows the reader to later see the similarity to the online techniques we develop. This section also provides some of the basic notation that we use in the sequel. Section 3 gives a first taste of the primal–dual approach in the context of online algorithms via the well-understood ski rental problem. We show an alternative way of deriving optimal algorithms for the ski rental problem using a simple primal–dual approach. In Section 4, we lay the foundations for the online primal–dual approach and design the basic algorithms for the packing–covering framework. We also study two toy examples that demonstrate the online framework. The rest of the sections show how to apply the primal–dual approach to many interesting and fundamental problems. We tried to make the chapters independent of each other; however, there are still certain connections between chapters, and thus closely related problems appear in consecutive chapters and typically in increasing order of complexity.

Among the problems that we consider are the weighted caching problem, generalized caching, the online set-cover problem, several graph optimization problems, routing, load balancing, and even the problem of allocating ad-auctions. We also show that classic online problems like the dynamic TCP-acknowledgement problem can be optimally solved using a primal–dual approach. There are also several more problems that can be solved via the primal–dual approach and are not discussed here. Such problems are, for example, the admission control problem [5], the parking permit problem [83] and the inventory problem [31].

References

[1] D. Achlioptas, M. Chrobak, and J. Noga, "Competitive analysis of randomized paging algorithms," *Theory Computer Science*, vol. 234, no. 1–2, pp. 203–218, 2000.

[2] S. Albers, S. Arora, and S. Khanna, "Page replacement for general caching problems," in *Proceedings of the 10th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 31–40, 1999.

[3] N. Alon, B. Awerbuch, Y. Azar, N. Buchbinder, and J. Naor, "The online set cover problem," in *Proceedings of the 35th Annual ACM Symposium on the Theory of Computation*, pp. 100–105, 2003. (To appear, *SIAM Journal on Computing*).

[4] N. Alon, B. Awerbuch, Y. Azar, N. Buchbinder, and J. Naor, "A general approach to online network optimization problems," *ACM Transactions on Algorithms*, vol. 2, no. 4, pp. 640–660, 2006.

[5] N. Alon, Y. Azar, and S. Gutner, "Admission control to minimize rejections and online set cover with repetitions," in *Proceedings of the 17th Annual ACM Symposium on Parallelism in Algorithms and Architectures*, pp. 238–244, 2005.

[6] N. Alon and J. H. Spencer, *The Probabilistic Method*. New York, Wiley, Second ed., 2000.

[7] N. Andelman and Y. Mansour, "Auctions with budget constraints," in *Proceedings of the 9th Scandinavian Workshop on Algorithm Theory*, pp. 26–38, 2004.

[8] J. Aspnes, Y. Azar, A. Fiat, S. A. Plotkin, and O. Waarts, "On-line routing of virtual circuits with applications to load balancing and machine scheduling," *Journal of the ACM*, vol. 44, no. 3, pp. 486–504, 1997.

[9] B. Awerbuch, Y. Azar, and Y. Bartal, "On-line generalized Steiner problem," in *Proceedings of the 7th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 68–74, 1996.

[10] B. Awerbuch, Y. Azar, A. Fiat, and T. Leighton, "Making commitments in the face of uncertainty: How to pick a winner almost every time," in *Proceedings of the 28th Annual ACM Symposium on Theory of Computing*, pp. 519–530, 1996.

[11] B. Awerbuch, Y. Azar, and S. Plotkin, "Throughput-competitive online routing," in *Proceedings of the 34th Symposium on Foundations of Computer Science*, pp. 32–40, 1993.

[12] Y. Azar, "On-line load balancing," *Online Algorithms — The State of the Art,* Springer, pp. 178–195, 1998.

[13] Y. Azar, J. Naor, and R. Rom, "The competitiveness of on-line assignments," *Journal of Algorithms*, vol. 18, pp. 221–237, 1995.

[14] N. Bansal, N. Buchbinder, and J. S. Naor, "A primal–dual randomized algorithm for weighted paging," in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science*, pp. 507–517, 2007.

[15] N. Bansal, B. Niv, and N. J. Seffi, "Randomized competitive algorithms for generalized caching," in *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pp. 235–244, 2008.

[16] A. Bar-Noy, R. Bar-Yehuda, A. Freund, J. S. Naor, and B. Schieber, "A unified approach to approximating resource allocation and scheduling," *Journal of the ACM*, vol. 48, no. 5, pp. 1069–1090, 2001.

[17] R. Bar-Yehuda and S. Even, "A linear-time approximation algorithm for the weighted vertex cover problem," *Journal of Algorithms*, vol. 2, pp. 198–203, 1981.

[18] Y. Barta, B. Bollobás, and M. Mendel, "A Ramsey-type theorem for metric spaces and its applications for metrical task systems and related problems," in *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, p. 396, 2001.

[19] Y. Bartal, A. Blum, C. Burch, and A. Tomkins, "A polylog(n)-competitive algorithm for metrical task systems," in *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, pp. 711–719, 1997.

[20] Y. Bartal, N. Linial, M. Mendel, and A. Naor, "On metric Ramsey-type phenomena," in *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pp. 463–472, 2003.

[21] Y. Bartal and M. Mendel, "Randomized k-server algorithms for growth-rate bounded graphs," *Journal of Algorithms*, vol. 55, no. 2, pp. 192–202, 2005.

[22] P. Berman and C. Coulston, "On-line algorithms for Steiner tree problems," in *Proceedings of the 29th Annual ACM Symposium on the Theory of Computation*, pp. 344–353, 1997.

[23] D. Bertsekas and R. Gallager, *Data Networks.* Prentice-Hall, Inc., 1987.

[24] A. Blum, C. Burch, and A. Kalai, "Finely-competitive paging," *IEEE Symposium on Foundations of Computer Science*, pp. 450–458, 1999.

[25] A. Blum, M. Furst, and A. Tomkins, "What to do with your free time: Algorithms for infrequent requests and randomized weighted caching," Manuscript, 1996.

[26] A. Blum and J. Hartline, “Near-optimal online auctions,” in *Proceedings of the 16th Annual ACM–SIAM Symposium on Discrete Algorithms*, 2005.

[27] C. Borgs, J. Chayes, N. Immorlica, M. Mahdian, and A. Saberi, “Multi-unit auctions with budget-constrained bidders,” in *Proceedings of the 6th ACM Conference on Electronic Commerce*, pp. 44–51, 2005.

[28] A. Borodin and R. El-Yaniv, *Online Computation and Competitive Analysis*. Cambridge University Press, 1998.

[29] A. Borodin, N. Linial, and M. E. Saks, “An optimal on-line algorithm for metrical task system,” *Journal of the ACM*, vol. 39, no. 4, pp. 745–763, 1992.

[30] N. Buchbinder, K. Jain, and J. S. Naor, “Online primal–dual algorithms for maximizing ad-auctions revenue,” in *Proceedings of the 15th Annual European Symposium*, pp. 253–264, 2007.

[31] N. Buchbinder, T. Kimbrel, R. Levi, K. Makarychev, and M. Sviridenko, “Online make-to-order joint replenishment model: Primal dual competitive algorithms,” in *Proceedings of the 19th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 952–961, 2008.

[32] N. Buchbinder and J. Naor, “Online primal–dual algorithms for covering and packing problems,” in *Proceedings of the 13th Annual European Symposium on Algorithms*, 2005.

[33] N. Buchbinder and J. Naor, “Improved bounds for online routing and packing via a primal–dual approach,” in *Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science* (*FOCS 2006*), 2006.

[34] N. Buchbinder and J. Naor, “Fair online load balancing,” in *Proceedings of the 18th ACM Symposium on Parallelism in Algorithms and Architectures*, 2006.

[35] P. Cao and S. Irani, “Cost-aware WWW proxy caching algorithms,” in *Proceedings of the 1997 Usenix Symposium on Internet Technologies and Systems* (*USITS-97*), 1997.

[36] R. D. Carr, L. K. Fleischer, V. J. Leung, and C. A. Phillips, “Strengthening integrality gaps for capacitated network design and covering problems,” in *Proceedings of the 11th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 106–115, 2000.

[37] C. Chekuri, G. Even, A. Gupta, and D. Segev, “Set connectivity problems in undirected graphs and the directed steiner network problem,” in *Proceedings of the 19th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 532–541, 2008.

[38] C. Chekuri, S. Khanna, J. Naor, and L. Zosin, “A linear programming formulation and approximation algorithms for the metric labeling problem,” *SIAM Journal on Discrete Mathematics*, vol. 18, no. 3, pp. 608–625, 2005.

[39] C. Chekuri, S. Khanna, and F. B. Shepherd, “The all-or-nothing multicommodity flow problem,” in *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pp. 156–165, 2004.

[40] M. Chrobak, H. Karloff, T. Payne, and S. Vishwanathan, “New results on server problems,” *SIAM Journal on Discrete Mathmatics*, vol. 4, no. 2, pp. 172–181, 1991.

[41] V. Chvátal, “A greedy heuristic for the set-covering problem,” *Mathematics of Operations Research*, pp. 233–235, 1979.

[42] V. Chvátal, *Linear Programming*. W.H. Freeman, 1983.
[43] E. Cohen and H. Kaplan, "LP-based analysis of greedy-dual-size," in *Proceedings of the 10th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 879–880, 1999.
[44] B. Csaba and S. Lodha, "A randomized on-line algorithm for the k-server problem on a line," *Random Structures and Algorithms*, vol. 29, no. 1, pp. 82–104, 2006.
[45] D. R. Dooly, S. Goldman, and S. D. Scott, "On-line analysis of the TCP acknowledgement delay problem," *Journal of the ACM*, vol. 48, pp. 243–273, 2001.
[46] J. Fakcharoenphol, S. Rao, and K. Talwar, "A tight bound on approximating arbitrary metrics by tree metrics," in *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, pp. 448–455, 2003.
[47] U. Feige, "A threshold of in for approximating set cover," *Journal of the ACM*, vol. 45, no. 4, pp. 634–652, 1998.
[48] A. Fiat, R. M. Karp, M. Luby, L. A. McGeoch, D. D. Sleator, and N. E. Young, "Competitive paging algorithms," *Journal of Algorithms*, vol. 12, no. 4, pp. 685–699, 1991.
[49] A. Fiat and M. Mendel, "Better algorithms for unfair metrical task systems and applications," *SIAM Journal on Computing*, vol. 32, no. 6, pp. 1403–1422, 2003.
[50] A. Fiat and M. Mendel, "Better algorithms for unfair metrical task systems and applications," *SIAM Journal on Computing*, vol. 32, no. 6, pp. 1403–1422, 2003.
[51] A. Fiat, Y. Rabani, and Y. Ravid, "Competitive k-server algorithms," *Journal of Computer and System Sciences*, vol. 48, no. 3, pp. 410–428, 1994.
[52] L. Fleischer, "A fast approximation scheme for fractional covering problems with variable upper bounds," in *Proceedings of the 15th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 1001–1010, 2004.
[53] L. K. Fleischer, "Approximating fractional multicommodity flow independent of the number of commodities," *SIAM Journal on Discrete Mathematics*, vol. 13, no. 4, pp. 505–520, 2000.
[54] N. Garg and R. Khandekar, "Fractional covering with upper bounds on the variables: Solving LPs with negative entries," in *Proceedings of the 12th European Symposium on Algorithms*, pp. 371–382, 2004.
[55] N. Garg and J. Konemann, "Faster and simpler algorithms for multicommodity flow and other fractional packing problems," in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, pp. 300–309, 1998.
[56] N. Garg, G. Konjevod, and R. Ravi, "A polylogarithmic approximation algorithm for the group steiner tree problem," *Journal of Algorithms*, vol. 37, no. 1, pp. 66–84, 2000.
[57] A. Goel and A. Meyerson, "Simultaneous optimization via approximate majorization for concave profits or convex costs," *Algorithmica*, vol. 44, no. 4, pp. 301–323, 2006.
[58] A. Goel, A. Meyerson, and S. A. Plotkin, "Approximate majorization and fair online load balancing," in *Proceedings of the 12th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 384–390, 2001.

[59] A. Goel, A. Meyerson, and S. A. Plotkin, "Combining fairness with throughput: online routing with multiple objectives," *Journal of Computer and System Science*, vol. 63, no. 1, pp. 62–79, 2001.

[60] M. Goemans and D. Williamson, "A general approximation technique for constrained forest problems," *SIAM Journal on Computing*, vol. 24, pp. 296–317, 1995.

[61] R. L. Graham, "Bounds for certain multiprocessing anomalies," *Bell System Technical Journal*, vol. 45, no. 1, pp. 1563–1581, 1966.

[62] D. Hochbaum, *Approximation Algorithms for NP-Hard Problems*. Boston, MA: PWS Publishing Company, 1997.

[63] M. Imase and B. Waxman, "Dynamic Steiner tree problem," *SIAM Journal Discrete Mathematics*, vol. 4, pp. 369–384, 1991.

[64] S. Irani, "Competitive analysis of paging: A survey," in *Proceedings of the Workshop on Online Algorithms*, Dagstuhl, Germany.

[65] S. Irani, "Page replacement with multi-size pages and applications to web caching," in *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, pp. 701–710, 1997.

[66] S. Irani, "Randomized weighted caching with two page weights," *Algorithmica*, vol. 32, no. 4, pp. 624–640, 2002.

[67] J. Jaffe, "Bottleneck flow control," *IEEE Transactions on Communications*, vol. 29, no. 7, pp. 954–962, 1981.

[68] D. S. Johnson, "Approximation algorithms for combinatorial problems," *Journal of Computer and System Sciences*, vol. 9, pp. 256–278, 1974.

[69] B. Kalyanasundaram and K. R. Pruhs, "An optimal deterministic algorithm for online b-matching," *Theoretical Computer Science*, vol. 233, no. 1–2, pp. 319–325, 2000.

[70] A. R. Karlin, C. Kenyon, and D. Randall, "Dynamic TCP acknowledgement and other stories about $e/(e-1)$," in *Proceedings of the ACM Symposium on Theory of Computing*, pp. 502–509, 2001.

[71] A. R. Karlin, M. S. Manasse, L. A. McGeoch, and S. S. Owicki, "Competitive randomized algorithms for nonuniform problems," *Algorithmica*, vol. 11, no. 6, pp. 542–571, 1994.

[72] A. R. Karlin, M. S. Manasse, L. Rudolph, and D. D. Sleator, "Competitive snoopy caching," in *Proceedings of the 27th Annual Symposium on Foundations of Computer Science*, pp. 244–254, 1986.

[73] R. Karp, U. Vazirani, and V. Vazirani, "An optimal algorithm for online bipartite matching," in *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing*, pp. 352–358, 1990.

[74] J. Kleinberg, E. Tardos, and Y. Rabani, "Fairness in routing and load balancing," in *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, p. 568, 1999.

[75] S. G. Kolliopoulos and N. E. Young, "Tight approximation results for general covering integer programs," in *Proceedings of the IEEE Symposium on Foundations of Computer Science*, pp. 522–528, 2001.

[76] E. Koutsoupias and C. H. Papadimitriou, "On the k-server conjecture," *Journal of the ACM*, vol. 42, no. 5, pp. 971–983, 1995.

[77] A. Kumar and J. M. Kleinberg, "Fairness measures for resource allocation," in *Proceedings of the IEEE Symposium on Foundations of Computer Science*, pp. 75–85, 2000.

[78] L. Lovász, "On the ratio of optimal and fractional covers," *Discrete Mathematics*, vol. 13, pp. 383–390, 1975.

[79] M. Luby and N. Nisan, "A parallel approximation algorithm for positive linear programming," in *Proceedings of the 25th Annual ACM Symposium on Theory of Computing*, pp. 448–457, 1993.

[80] M. Mahdian and A. Saberi, "Multi-unit auctions with unknown supply," in *Proceedings of the 7th ACM Conference on Electronic Commerce*, pp. 243–249, 2006.

[81] L. A. McGeoch and D. D. Sleator, "A strongly competitive randomized paging algorithm," *Algorithmica*, vol. 6, no. 6, pp. 816–825, 1991.

[82] A. Mehta, A. Saberi, U. Vazirani, and V. Vazirani, "Adwords and generalized online matching," *Journal of the ACM*, vol. 54, no. 5, p. 22, 2007.

[83] A. Meyerson, "The parking permit problem," in *Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pp. 274–284, 2005.

[84] S. Plotkin, D. Shmoys, and E. Tardos, "Fast approximation algorithms for fractional packing and covering problems," *Mathematics of Operations Research*, vol. 20, pp. 257–301, 1995.

[85] P. Raghavan, "Probabilistic construction of deterministic algorithms: Approximating packing integer programs," *Journal of Computer and System Sciences*, vol. 37, no. 2, pp. 130–143, 1988.

[86] P. Raghavan and C. Thompson, "Randomized rounding: A technique for provably good algorithms and algorithmic proofs," *Combinatorica*, vol. 7, pp. 365–374, 1987.

[87] S. S. Seiden, "A general decomposition theorem for the k-server problem," in *Proceedings of the 9th Annual European Symposium on Algorithms*, pp. 86–97, 2001.

[88] D. D. Sleator and R. E. Tarjan, "Amortized efficiency of list update and paging rules," *Communications of the ACM*, vol. 28, no. 2, pp. 202–208, 1985.

[89] A. Srinivasan, "Improved approximation guarantees for packing and covering integer programs," *SIAM Journal on Computing*, vol. 29, no. 2, pp. 648–670, 1999.

[90] V. V. Vazirani, *Approximation Algorithms.* Springer, 2001.

[91] N. Young, "On-line caching as cache size varies," in *SODA '91: Proceedings of the 2nd Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 241–250, 1991.

[92] N. E. Young, "The k-server dual and loose competitiveness for paging," *Algorithmica*, vol. 11, no. 6, pp. 525–541, 1994.

[93] N. E. Young, "Randomized rounding without solving the linear program," in *Proceedings of the 6th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 170–178, 1995.

[94] N. E. Young, "On-line file caching," in *Proceedings of the 9th Annual ACM–SIAM Symposium on Discrete Algorithms*, pp. 82–86, 1998.