

# Αναφορά για Online Learning for Min Sum Set Cover and Pandora's Box

Kasionis Ioannis  
University of Piraeus  
ioannis.kasionis@gmail.com

Triantafyllos Petros  
University of Piraeus  
petrostriantafyllos@outlook.com

January 2025

## Abstract

Αυτή η αναφορά παρουσιάζει μια ανάλυση και περίληψη του άρθρου *Online Learning for Min Sum Set Cover and Pandora's Box*, το οποίο προσφέρει μια απλή αλλά αποτελεσματική μεθοδολογία για τη δημιουργία online learning αλγορίθμων για το *Pandora's Box Problem*, το *Min Sum Set Cover Problem (MSSC)* και άλλα συναφή προβλήματα στο πεδίο της *Stochastic optimization* [1]. Παραδοσιακά, αυτά τα προβλήματα υποθέτουν πρότερη γνώση των κατανομών τιμών. Παρ' όλα αυτά, σε αυτήν την εργασία μελετάται μια online learning προσέγγιση, όπου παρουσιάζονται πληροφορίες σε πολλούς γύρους.

Επιπροσθέτως, η έρευνα επεκτείνει αυτό το πλαίσιο σε bandit setting, όπου μόνο οι τιμές των boxes που έχουν ανοιχτεί αποκαλύπτονται στον learner μετά από κάθε γύρο. Αυτή η επέκταση καλύπτει και εναλλακτικές εκδοχές των προαναφερθέντων προβλημάτων [1].

## 1 Εισαγωγή

Τα δύο βασικά προβλήματα στη *Stochastic Optimization* είναι το *Pandora's Box* και το *MSSC*. Το *Pandora's Box*, που εισήχθη από τον Weitzman [2], παρέχει ένα γενικό πλαίσιο για τη βελτιστοποίηση αποφάσεων υπό αβεβαιότητα, με στόχο την ελαχιστοποίηση του κόστους από την αποκάλυψη πληροφοριών. Το *MSSC*, μια παραλλαγή του *Pandora's Box*, ασχολείται με περιπτώσεις όπου οι τιμές είναι είτε 0 είτε  $\infty$  [3].

Ενώ η παραδοσιακή διατύπωση των προβλημάτων αυτών υποθέτει πρόσβαση σε γνωστές στοχαστικές κατανομές [1], η συγκεκριμένη μελέτη επεκτείνει αυτά τα πλαίσια σε *online* σενάρια, όπου τα δεδομένα φτάνουν σειριακά χωρίς πρότερη διανομή.

Η οπτική γωνία του *online learning* φέρνει στο προσκήνιο προκλήσεις αλλά και ευκαιρίες. Σε αυτό το μοντέλο, παρουσιάζονται είτε αντιπαραθετικά (adversarial) είτε ανεξάρτητα σενάρια σε κάθε γύρο, και ο αλγόριθμος οφείλει να αποφασίζει την επόμενη ενέργεια βάσει των τωρινών δεδομένων και των προηγούμενων παρατηρήσεων. Η απόδοση μετριέται συνήθως μέσω *regret*, ο οποίος υπολογίζει το σωρευτικό κόστος του αλγορίθμου συγκριτικά με μια βέλτιστη offline πολιτική [4].

Τα βασικά συμβολή της παρούσας εργασίας συνοψίζονται ως εξής:

1. Προτείνεται ένας αποδοτικός online αλγόριθμος που επιτυγχάνει constant-competitive εγγυήσεις τόσο για το MSSC όσο και για τις γενικεύσεις του Pandora's Box σε full-information συνθήκες. Σε αυτό περιλαμβάνονται προβλήματα με σύνθετους περιορισμούς, όπως matroid bases [5].

2. Επεκτείνεται η ανάλυση σε *bandit settings*, όπου διατίθεται μόνο μερική πληροφόρηση (partial feedback) για τα επιλεγμένα boxes [6]. Παρά αυτήν τη μειωμένη πληροφόρηση, ο αλγόριθμός μας διατηρεί λογαριθμικό regret, προσφέροντας ισχυρές εγγυήσεις απόδοσης.
3. Με τη χρήση convex relaxations και τεχνικών online convex optimization, παρουσιάζονται πρακτικές προσεγγίσεις που γεφυρώνουν το θεωρητικό πλαίσιο με την υπολογιστική αποδοτικότητα [7].

Οι συνέπειες των παραπάνω αποτελεσμάτων είναι διττές. Πρώτον, διευρύνουν την εφαρμοσιμότητα του MSSC και του Pandora's Box σε δυναμικές, πραγματικές καταστάσεις, όπου οι αποφάσεις προσαρμόζονται σταδιακά σε νέα δεδομένα. Δεύτερον, η ενσωμάτωση τεχνικών online convex optimization αναδεικνύει τον κρίσιμο ρόλο των μαθηματικών εργαλείων στην αντιμετώπιση συνδυαστικών προβλημάτων [1].

Η υπόλοιπη εργασία δομείται ως ακολούθως: Στο Κεφάλαιο 2 παρουσιάζονται οι διατυπώσεις των προβλημάτων και η σχετική βιβλιογραφία, στο Κεφάλαιο 3 περιγράφονται οι προτεινόμενες μέθοδοι, και στα Κεφάλαια 4 και 5 ακολουθούν η θεωρητική ανάλυση και τα πειραματικά αποτελέσματα, αντίστοιχα. Τέλος, παρατίθενται συμπεράσματα και μελλοντικές επεκτάσεις.

## 2 Ορισμοί Προβλημάτων

### 2.1 Pandora's Box Problem

Το *Pandora's Box Problem* περιλαμβάνει ένα σύνολο από  $n$  boxes, καθένα από τα οποία περιέχει μια άγνωστη τιμή και ένα κόστος για να ανοιχτεί. Ο στόχος είναι να βρεθεί η σειρά ανοίγματος των boxes ώστε να ελαχιστοποιείται το συνολικό κόστος επιθεώρησης μαζί με την τιμή του επιλεγμένου αντικειμένου. Ειδικότερα, δεδομένης μιας κατανομής πιθανών τιμών και κόστους για κάθε box, σκοπός είναι να προσδιοριστεί μια στρατηγική που εξισορροπεί την αναζήτηση (άνοιγμα boxes) με την αξιοποίηση (επιλογή της βέλτιστης τιμής).

Στην online εκδοχή που εξετάζεται σε αυτή την εργασία, οι τιμές και τα κόστη των boxes αποκαλύπτονται αντιπαραθετικά σε  $T$  γύρους, απαιτώντας από τον learner να υιοθετεί στρατηγικές που προσαρμόζονται σε νέα δεδομένα, ελαχιστοποιώντας παράλληλα το regret έναντι μιας βέλτιστης offline πολιτικής.

### 2.2 Min Sum Set Cover (MSSC)

Το *MSSC* αποτελεί μια ειδική περίπτωση του *Pandora's Box*, όπου οι τιμές (ή τα κόστη) στα boxes είναι περιορισμένα στις τιμές 0 ή  $\infty$ . Το *MSSC* είναι ιδιαίτερα χρήσιμο σε εφαρμογές όπου το αποτέλεσμα είναι δυαδικό, υποδεικνύοντας αν ένα αντικείμενο πληροί ή όχι συγκεκριμένα κριτήρια.

## 3 Μεθοδολογίες και Αποτελέσματα

Οι συγγραφείς ανέπτυξαν ένα τριμερές πλαίσιο, εμπνευσμένο από τα scenario-aware relaxations του [?], προσαρμοσμένο όμως σε online περιβάλλον. Το πλαίσιο εφαρμόζεται σε δύο benchmarks—*non-adaptive (NA)* και *partially adaptive (PA)*—υπό full-information και bandit feedback.

### Scenario-Aware Relaxation και Rounding

Αρχικά, σε κάθε γύρο διατυπώνεται το πρόβλημα (επιλογή ενός box, επιλογή  $k$  boxes ή επιλογή matroid basis) ως ένα *scenario-aware* κυρτό πρόγραμμα, του οποίου η κλασματική λύση αναπαριστά την ιδέα του «ανοίγματος» ή «επιλογής» boxes. Με βάση το [?], αυτή η χαλάρωση σχεδιάζεται έτσι ώστε ένας κοινός αλγόριθμος rounding να μπορεί να εφαρμοστεί *ανεξαρτήτως του σεναρίου* που αποκαλύπτεται. Στην πράξη, μια  $\alpha$ -approximate διαδικασία rounding εγγυάται ότι η μετάβαση

από τη κλασματική στη διακριτή λύση (δηλαδή την πραγματική επιλογή boxes ή μια σειρά ανοίγματος) αυξάνει το κόστος κατά έναν σταθερό παράγοντα  $\alpha$ . Αυτή η ιδέα αποτρέπει την εκ νέου βελτιστοποίηση για κάθε σενάριο και είναι κρίσιμη για τη διατήρηση no-regret εγγυήσεων.

## Online Convex Optimization για Full Information vs. PA

Για την αντιμετώπιση ενός *partially adaptive* benchmark σε *full-information* περιβάλλον, οι συγγραφείς χρησιμοποιούν *Follow the Regularized Leader (FTRL)*. Σε κάθε γύρο, αφού παρατηρήσουν τη συνάρτηση κόστους για το χαλαρό πρόβλημα, ανανεώνουν την κλασματική τους απόφαση ελαχιστοποιώντας το άθροισμα των προηγούμενων απωλειών και ενός ισχυρά κυρτού ρυθμιστή. Η αντικατάσταση των μεθόδων gradient-descent με το FTRL επιτρέπει αυστηρότερους bound στο regret. Στη συνέχεια, μόλις επιλεγεί η κλασματική λύση, γίνεται rounding μέσω της ίδιας scenario-independent  $\alpha$ -approximation διαδικασίας. Στην ειδική περίπτωση του MSSC, αυτή η μέθοδος αποδίδει την *tight 4-approximation* που είναι γνωστή offline, αποδεικνύοντας ότι η προσέγγιση ταιριάζει με τα βέλτιστα γνωστά αποτελέσματα ακόμη και με αντιπαραθετική εισαγωγή δεδομένων.

Επιπλέον, ενάντια σε ένα partially adaptive benchmark, οι τεχνικές αυτές παρέχουν  $\alpha$ -approximate no-regret για προβλήματα τύπου Pandora's Box με διάφορους περιορισμούς. Στο άρθρο αναφέρονται:

$$\begin{cases} 9.22\text{-approx. no-regret για επιλογή 1 box,} \\ O(1)\text{-approx. no-regret για επιλογή } k \text{ boxes,} \\ O(\log k)\text{-approx. no-regret για επιλογή matroid basis.} \end{cases}$$

Αυτά τα αποτελέσματα ισχύουν στο full-information μοντέλο, αποδεικνύοντας σταθερές ή λογαριθμικές αποκλίσεις σε σχέση με μια βέλτιστη εκ των υστέρων (in hindsight) partially adaptive στρατηγική.

## Bandit vs. PA μέσω Μίξης FTRL και «Open-All»

Σε *bandit feedback* πλαίσιο, ο learner συνήθως παρατηρεί μόνο τη συνάρτηση κόστους  $f^{(t)}(x_t)$  στο επιλεγμένο σημείο  $x_t$ , χωρίς να έχει πρόσβαση σε όλη τη συνάρτηση κάθε γύρο. Για να αντισταθμιστεί η έλλειψη πληροφόρησης, ο αλγόριθμος των συγγραφέων εναλλάσσεται τυχαία ανάμεσα σε:

1. *Άνοιγμα όλων των boxes* (πληρώνοντας ένα upfront κόστος), ώστε να μαθαίνει πλήρως τη συνάρτηση κόστους για αυτό τον γύρο.
2. *Βήμα OCO* (παρόμοιο με FTRL) χρησιμοποιώντας μόνο τη μερική πληροφόρηση από τα boxes που επιλέχθηκαν στους άλλους γύρους.

Αυτή η στρατηγική ισορροπεί ανάμεσα στην εξερεύνηση και την εκμετάλλευση, διατηρώντας τον ίδιο  $\alpha$ -approximation παράγοντα στο rounding, και εξασφαλίζοντας *sublinear* regret έναντι του PA benchmark. Διαισθητικά, αν δεν ανοίξουμε ποτέ όλα τα boxes, οι εκτιμήσεις για τα gradient θα είναι κακές, ενώ αν τα ανοίγουμε πάντα, ξοδεύουμε υπερβολικά για επιθεώρηση. Η ισορροπία μεταξύ αυτών των άκρων οδηγεί σε ικανοποιητικές επιδόσεις με σταθερούς ή  $O(\log k)$  παράγοντες προσέγγισης, όπως και στην περίπτωση του full-information.

## Bandit vs. NA και ο ρόλος των μη-Lipschitz Costs

Ένα *non-adaptive* benchmark καθορίζει ένα υποσύνολο boxes για όλη τη διάρκεια των γύρων. Η συνάρτηση κόστους σε αυτές τις περιπτώσεις είναι συχνά *μη-Lipschitz*, γεγονός που δυσχεραίνει την άμεση χρήση του FTRL. Επομένως, στο άρθρο παρουσιάζεται μια διαδικασία *explore-exploit* στην οποία:

1. **Explore:** Ο αλγόριθμος ανοίγει πλήρως τα boxes σε ορισμένους γύρους, ανακαλύπτοντας ένα νέο constraint (π.χ. ένα threshold ή μια περιοχή τιμών). Έπειτα, το ενσωματώνει σε ένα παγκόσμιο γραμμικό πρόγραμμα και χρησιμοποιεί την ελλειψοειδή μέθοδο (ellipsoid subroutine) για να βρει μια εφικτή κλασματική λύση.
2. **Exploit:** Αυτή η κλασματική λύση μετατρέπεται σε non-adaptive σύνολο boxes, το οποίο χρησιμοποιείται μέχρι τον επόμενο γύρο explore.

**Το Θεώρημα 5.3** στο άρθρο αναφέρει ότι αν υπάρχει μερικώς προσαρμόσιμη (partially adaptive) στρατηγική που είναι  $\beta$ -competitive σε σχέση με τη NA βέλτιστη πολιτική για κάποιο πρόβλημα (επιλογή ενός box,  $k$  boxes ή matroid), τότε ο παρουσιαζόμενος bandit αλγόριθμος ταιριάζει το  $\beta$  μέχρι έναν μικρό σταθερό παράγοντα, συν έναν  $o(T)$  όρο στο regret. Χρησιμοποιώντας ήδη γνωστές σταθερές για αυτά τα προβλήματα, προκύπτουν προσεγγιστικές εγγυήσεις no-regret:

$$\begin{cases} 3.16\text{-approx. no-regret για επιλογή 1 box,} \\ 12.64\text{-approx. no-regret για επιλογή } k \text{ boxes,} \\ O(\log k)\text{-approx. no-regret για επιλογή matroid basis.} \end{cases}$$

Έτσι, ακόμη και στο πιο περιοριστικό NA πλαίσιο και υπό μερική πληροφόρηση, η προσέγγιση παραμένει σχεδόν βέλτιστη.

## Συμπεράσματα

Συνολικά, η μεθοδολογία αυτή—scenario-aware convex relaxations, online convex optimization για πλήρη ή μερική πληροφόρηση, και scenario-independent rounding—προσφέρει ενιαίες *constant-competitive no-regret* λύσεις ακόμη και σε αντιπαραθετικές εκδοχές των Pandora’s Box και MSSC. Παρότι η βασική ιδέα παραμένει κοινή για PA και NA benchmarks, οι *σταθεροί παράγοντες* διαφέρουν λόγω της διαφορετικής φύσης κάθε μοντέλου. Εντέλει, αποδεικνύεται ότι ακόμη και σε bandit περιβάλλοντα, είναι εφικτή η επίτευξη απόδοσης συγκρίσιμης με τις offline προσεγγίσεις, συμπεριλαμβανομένης της βέλτιστης 4-approximation για το MSSC, γεγονός που αναδεικνύει τη δυναμική της σύζευξης convex relaxations, online learning και αποδοτικών rounding τεχνικών.

## 4 Σύγκριση με Προηγούμενη Εργασία

Η παρούσα εργασία βασίζεται σε θεμελιώδη αποτελέσματα στη στοχαστική βελτιστοποίηση, ειδικά στους τομείς του Pandora’s Box και του MSSC. Προηγούμενες μελέτες επικεντρώθηκαν στην στοχαστική περίπτωση, όπου οι τιμές και τα κόστη προέρχονται από γνωστές κατανομές, επιτρέποντας αποδοτικούς offline αλγόριθμους και μερικώς προσαρμόσιμες στρατηγικές [2, 8].

Αντίθετα, η δική μας έρευνα στρέφεται στο online περιβάλλον, όπου τιμές και κόστη ορίζονται αντιπαραθετικά σε πολλούς γύρους. Αξίζει να σημειωθεί ότι το προτεινόμενο πλαίσιο διευκολύνει σημαντικά την ανάπτυξη ανταγωνιστικών αλγορίθμων μέσω convex relaxations και τεχνικών online convex optimization, επεκτείνοντας προηγούμενες εργασίες, ιδίως όσον αφορά τις εγγυήσεις σε bandit συνθήκες [1]. Επιπλέον, τα αποτελέσματά μας γενικεύουν αλγορίθμους του MSSC επιτυγχάνοντας constant-factor προσεγγίσεις ακόμη και υπό matroid περιορισμούς, μια πρόκληση που παλαιότερες εργασίες είχαν δυσκολία να αντιμετωπίσουν αποτελεσματικά [3].

## 5 Εφαρμογές και Μελλοντική Εργασία

Τα συμπεράσματα αυτής της μελέτης βρίσκουν πρακτική εφαρμογή σε ποικίλους τομείς, μεταξύ άλλων:

- **Διαχείριση Πόρων:** Βελτιστοποίηση του κόστους επιθεώρησης σε βιομηχανικές διαδικασίες ή συστήματα ποιοτικού ελέγχου.

- **Αλγόριθμοι Αναζήτησης:** Βελτίωση στρατηγικών για την εύρεση βέλτιστων λύσεων υπό αβεβαιότητα.
- **Online Δημοπρασίες:** Προσαρμογή στρατηγικών πλειοδοσίας σε αντιπαραθετικά περιβάλλοντα.

Μελλοντικές επεκτάσεις θα μπορούσαν να εξετάσουν δυναμικές παραλλαγές, όπου οι περιορισμοί μεταβάλλονται με την πάροδο του χρόνου, ή την ενσωμάτωση πιο πλούσιων μηχανισμών ανάδρασης στο bandit setting. Επιπλέον, παραμένει ζητούμενο η περαιτέρω βελτίωση των παραγόντων προσέγγισης υπό πιο σύνθετους συνδυαστικούς περιορισμούς, θέμα με σημαντική θεωρητική και πρακτική αξία.

## 6 Συμπέρασμα

Παρουσιάσαμε ένα ολοκληρωμένο πλαίσιο για την αντιμετώπιση των online εκδοχών των Pandora's Box και Min Sum Set Cover (MSSC), δύο θεμελιωδών προβλημάτων στη στοχαστική βελτιστοποίηση. Αξιοποιώντας convex relaxations και τεχνικές online convex optimization, οι συγγραφείς ανέπτυξαν αποδοτικούς αλγορίθμους που επιτυγχάνουν constant competitiveness σε full information και bandit περιβάλλοντα. Αυτοί οι αλγόριθμοι επεκτείνουν προηγούμενες προσεγγίσεις, ιδίως σε αντιπαραθετικά δεδομένα και σε πολύπλοκους περιορισμούς (π.χ. matroids) [1].

## References

- [1] Evangelia Gergatsouli and Christos Tzamos. Online learning for min sum set cover and pandora's box. In *International Conference on Machine Learning*, pages 7382–7403. PMLR, 2022.
- [2] Martin Weitzman. *Optimal search for the best alternative*, volume 78. Department of Energy, 1978.
- [3] Uriel Feige, László Lovász, and Prasad Tetali. Approximating min sum set cover. *Algorithmica*, 40:219–234, 2004.
- [4] Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012.
- [5] Nikhil Bansal, Anupam Gupta, and Ravishankar Krishnaswamy. A constant factor approximation algorithm for generalized min-sum set cover. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 1539–1545. SIAM, 2010.
- [6] Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. *arXiv preprint cs/0408007*, 2004.
- [7] Shai Shalev-Shwartz and Yoram Singer. A primal-dual perspective of online learning algorithms. *Machine Learning*, 69:115–142, 2007.
- [8] Shuchi Chawla, Evangelia Gergatsouli, Yifeng Teng, Christos Tzamos, and Ruimin Zhang. Pandora's box with correlations: Learning and approximation. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1214–1225. IEEE, 2020.