Minimax Theory and its Applications
Volume 12 (2027), No. 1, 29 to 33

Non-smooth minimax optimization of adaptive resource allocation under coupled uncertainty constraints

Dr. Anupa Sinha[1], Dr. Sanjay Kumar[2]

[1]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India.
ku.anupasinha@kalingauniversity.ac.in, https://orcid.org/0009-0009-4725-7923
[2]Assistant Professor, Kalinga University, Naya Raipur, Chhattisgarh, India.
ku.sanjaykumar@kalingauniversity.ac.in, https://orcid.org/0009-0004-2958-2902

Abstract:
A planner budgets n services, and wants to allocate this budget knowing only that the worst-case penalty for each shortfall is uncertain. The uncertainties are coupled: a budget Γ of size constrains the number of services that can be at their worst at a time. Selecting the allocation that minimises the worst-case penalty is a convex-concave minimax problem. It is non-smooth twice over: the shortfall functions have kinks and the opponent's best response swaps rank whenever two services tie. We derive the subdifferential of the worst case cost, provide a reference exact linear-programming reformulation, and compare three first-order solvers: a Danskin subgradient method, accelerated gradient on a fixed Nesterov smoothing, and this method with an adaptively shrinking smoothing parameter. On a 50-service instance the adaptive method obtains a relative gap of 5.0×10^(-4) in 2,000 iterations. The subgradient method gets 2.9×10^(-3), fixed smoothing stagnates at a parameter-dependent 'floor' between 10^(-3) and 10^(-1). The robust allocation itself has a distinctive signature: more services are partly short (17 rather than 9) so that each shortfall is too small to be worth attacking, and the worst-case penalty is reduced, for a modest-Cost sensitivity of 4.7%, by 2.6% at a worst-case penalty of 22.2%.

Keywords*:* non-smooth optimization; minimax; robust resource allocation; budgeted uncertainty; Nesterov smoothing; Danskin's theorem
MSC 2020: 90C47, 90C25, 49J52, 90B50

1. Introduction

Linear optimisation under uncertainty is typically presented as a continous problem. The archetype in the resource allocation literature: network utility maximisation (Kelly et al., 1998; Palomar & Chiang, 2006): concave utility functions, a budget, and clearing prices. Practice is not so clean. A hospital network allocating nursing hours, a cloud platform allocating servers, a charitable organisation allocating supplies encounter piecewise linear costs: they need not pay a cost while meeting demand but linear in what they fall short. The significance of missing these shortfalls is also uncertain, and these uncertainties are coupled. A natural disaster might render a handful of services critical, but probably not all.

The budgeted uncertainty set of Bertsimas and Sim (2004) models this coupling with one parameter Γ: the number of components that may be inflated simultaneously. Together with a linearization of the shortfall costs, this yields a minimax problem that is convex in the allocations and concave (indeed linear) in the uncertainty, but non-smooth on both sides.

Questions. We pose three questions with regards to the problem.

1. What is the subdifferential of the worst-case cost and what does optimal allocation mean?
2. Which first order method should be used if the problem needs to be re-solved many times, such as each planning period as forecasts are updated?
3. What does robustness buy and what does it cost?

Results in brief. (i) The worst-case cost is the nominal cost plus the sum of the Γ largest weighted shortfalls. Its subdifferential is given by Danskin (1966), its kinks come from ties in that ranking as well as from the shortfall functions. (ii) Nesterov smoothing (2005) gets stuck on a wall that rises with the smoothing parameter. We let the parameter shrink with the iteration count, which lifts the wall and beats the subgradient method by a factor of six at 2,000 iterations. (iii) Robust allocations are not where all come just short. They are where some services are marginally below the line, so the adversary does not get to hit the large single shortfalls that the adversary loves.

Section 2 states the model. Section 3 discusses its non-smooth structure. Section 4 shows the algorithms, Section 5 the experiments, and Section 6 discusses the limitations.

2. Model

Allocation. There are n services. Service i has nominal demand $\bar{d}_i > 0$ and converts one unit of resource into $s_i > 0$ units of service. A total budget B must be spent, and service i can absorb at most $\bar{x}_i$. The feasible set is

$$X = \left\{ x \in \mathbb{R}^n :\ 0 \le x_i \le \bar{x}_i,\ \sum_{i=1}^{n} x_i = B \right\}.$$

The unmet demand of service i is $h_i(x_i) = \left(\bar{d}_i - s_i x_i\right)_+$, which is convex, piecewise linear, and kinked at $x_i = \bar{d}_i / s_i$.

Coupled uncertainty. The penalty per unit of unmet demand is $u_i = \bar{u}_i(1 + \rho z_i)$. Here $\bar{u}_i > 0$ is the nominal penalty, $\rho > 0$ is the maximal relative increase, and z lies in the budgeted set

$$Z_\Gamma = \left\{ z \in [0,1]^n :\ \sum_{i=1}^{n} z_i \le \Gamma \right\}, \qquad 0 \le \Gamma \le n. \qquad (1)$$

The constraint $\sum_i z_i \le \Gamma$ is what couples the uncertainties. Any Γ services may be at their worst, but not all of them together. $\Gamma = 0$ is the nominal problem and $\Gamma = n$ is the fully pessimistic box model (Ben-Tal & Nemirovski, 1998).

Problem. The planner solves

$$\varphi^* = \min_{x \in X} \varphi(x), \qquad \varphi(x) = \max_{z \in Z_\Gamma} L(x,z), \qquad L(x,z) = \sum_{i=1}^{n} \bar{u}_i \,(1 + \rho z_i)\, h_i(x_i). \qquad (2)$$

L is convex in x, because each h_i is convex and the weights are positive, and linear in z. Both X and Z_Γ are compact and convex, so Sion's theorem gives min–max = max–min, and a saddle point (x^*, z^*) exists.

Adaptivity. In operation, (2) is not solved once. At each planning epoch t the forecasts $\bar{d}^{\,t}$, and possibly $\bar{u}^{\,t}$, are updated, and (2) is solved again from the previous allocation. This is the sense in which the allocation is adaptive. It puts a premium on solvers that are cheap per iteration, can be warm-started, and do not need to be retuned at every epoch. The experiments below study one epoch; Section 6 returns to the rolling setting.

3. Non-smooth structure

Write $a_i(x) = \bar{u}_i h_i(x_i) \ge 0$ for the nominal penalty of service i, and $a_{[1]} \ge a_{[2]} \ge \cdots$ for the sorted values.

Proposition 1 (closed form). For integer Γ,

$$\varphi(x) = \sum_{i=1}^{n} a_i\,(x) + \rho \sum_{k=1}^{\Gamma} a_{[k]}\,(x).$$

For fractional Γ the last term gains $(\Gamma - \lfloor \Gamma \rfloor)\, \rho\, a_{[\lfloor \Gamma \rfloor + 1]}$.

Proof. For fixed x, the inner problem maximises $\sum_i \rho\, a_i z_i$ over Z_Γ. This is a continuous knapsack with unit weights and nonnegative values, so the greedy choice $z_i = 1$ on the Γ largest a_i is optimal. ▫

The adversary's best response is therefore cheap, costing $O(n \log n)$. It is also discontinuous in x, because the set of the Γ largest values changes whenever two services swap rank.

Proposition 2 (subdifferential). φ is convex and finite on X. Let $Z^*(x)$ be the set of maximisers in (2), and for each i let

$$\Theta_i(x) = \begin{cases} \{1\}, & x_i < \bar{d}_i / s_i, \\ [0,1], & x_i = \bar{d}_i / s_i, \\ \{0\}, & x_i > \bar{d}_i / s_i. \end{cases}$$

Then

$$\partial \varphi(x) = \operatorname{conv}\{ (-\bar{u}_i(1 + \rho z_i)\, s_i\, \theta_i)_{i=1}^{n} :\ z \in Z^*(x),\ \theta_i \in \Theta_i(x) \}. \qquad (3)$$

Proof. φ is a pointwise maximum, over the compact set Z_Γ, of functions that are convex in x and continuous in (x, z). Danskin's theorem (1966), in its non-smooth form (Clarke, 1990, Sect. 2.8), gives $\partial \varphi(x) = \operatorname{conv} \bigcup_{z \in Z^*(x)} \partial_x L(x,z)$. For fixed z, L is separable in x, and $\partial h_i(x_i) = -s_i\, \Theta_i(x_i)$. ▫

Formula (3) shows the two sources of non-smoothness.

• **Kinks of type A (service).** Some x_i sits exactly at $\bar{d}_i / s_i$, so the service is just fully served.

• **Kinks of type B (ranking).** The Γ-th and $(\Gamma + 1)$-th largest a_i are tied, so $Z^*(x)$ is not a singleton.

Type B kinks are not incidental. They are where the robust optimum tends to lie, as the next result explains.

Proposition 3 (shadow-price optimality). $x^* \in X$ is optimal if and only if there exist $g \in \partial \varphi(x^*)$ and a scalar ν (the price of one unit of budget) such that, for every i,

$$g_i + \nu \begin{cases} \ge 0, & x_i^* = 0, \\ = 0, & 0 < x_i^* < \bar{x}_i, \\ \le 0, & x_i^* = \bar{x}_i. \end{cases}$$

Proof. This is the condition $0 \in \partial \varphi(x^*) + N_X(x^*)$. The normal cone of X at x^* consists of vectors $\nu \mathbf{1}$ plus sign-constrained terms on the active bounds. ▫

Interpretation. For a service that is partly served ($0 < x_i^* < \bar{d}_i / s_i$), Proposition 3 reads $\bar{u}_i(1 + \rho z_i) s_i = \nu$ for some worst-case z. Each partly served service must offer the same *worst-case* marginal value per unit of resource. In the nominal problem ($\Gamma = 0$) the values $\bar{u}_i s_i$ are generic and almost never equal. The solution therefore fully serves

the services with the highest $\bar{u}_i s_i$ and leaves the rest short, with at most one service partly served: a bang-bang allocation. In the robust problem the adversary's choice of z depends on the allocation. The planner can then equalise worst-case marginal values by bringing several shortfalls a_i to a common level, at which the adversary is indifferent between them. The optimum therefore sits on a type B kink, and the allocation spreads scarcity over more services. Section 5 confirms this numerically.

Proposition 4 (exact LP form). By linear-programming duality applied to the inner problem, as in Bertsimas and Sim (2004),

$$\varphi^* = \min_{x\in X,\, v,\pi,p} \left\{ \sum_i \bar{u}_i\, v_i + \Gamma\pi + \sum_i p_i \;:\; v_i \geq \bar{d}_i - s_i x_i,\ \pi + p_i \geq \rho \bar{u}_i v_i,\ v, p \geq 0,\ \pi \geq 0 \right\}. \qquad (4)$$

This LP has $3n + 1$ variables. We use it only to certify φ^* in the experiments. In the adaptive setting, first-order methods are preferable, because they warm-start naturally and their per-iteration cost is $O(n \log n)$.

4. Algorithms

All three methods use the same projection onto X. It is computed by bisection on the multiplier of $\sum_i x_i = B$, which needs $O(n)$ work per bisection step. Let R be a bound on the diameter of X.

4.1. Danskin subgradient method (SG)

At x_k, compute $z_k \in Z^*(x_k)$ greedily (Proposition 1) and take g_k from (3) with $\theta_i = \mathbf{1}\{h_i(x_i) > 0\}$. Update

$$x_{k+1} = \Pi_X \left(x_k - \frac{R}{\| g_k \| \sqrt{k}}\, g_k \right),$$

and keep the best value seen. For convex Lipschitz φ this gives $\varphi_{\text{best}} - \varphi^* = O\big(RG/\sqrt{k}\big)$, where G bounds $\| g \|$. The rate cannot be improved for general non-smooth convex functions with a subgradient oracle. Davis and Drusvyatskiy (2019) extend guarantees of this kind to weakly convex problems, which matters if the h_i are replaced by non-convex service curves.

4.2. Smoothing

Both sources of kinks are maxima of linear functions, so both can be smoothed with Nesterov's technique (2005). We use

$$H_\mu(t) = \max_{0\leq y\leq 1} \left\{ yt - \frac{\mu}{2} y^2 \right\}, \qquad \varphi_\mu(x) = \max_{z\in Z_\Gamma} \left\{ \sum_i \bar{u}_i\, (1 + \rho z_i)\, H_\mu\big(\bar{d}_i - s_i x_i\big) - \frac{\mu}{2} \| z \|^2 \right\}. \qquad (5)$$

H_μ is the Huber function, a smoothed version of $(t)_+$. The inner maximiser is $z_\mu(x) = \Pi_{Z_\Gamma}\big(\rho\, \bar{u} \circ H_\mu(\cdot)/\mu\big)$, a projection onto the capped simplex that is again computed by bisection.

Lemma 1. (a) The function φ_μ is differentiable with Lipschitz gradient, with constant $L_\mu = O(1/\mu)$. (b) It satisfies $0 \leq \varphi(x) - \varphi_\mu(x) \leq \mu D$ with $D = \Gamma/2 + (1 + \rho) \sum_i \bar{u}_i /2$.

Proof. For (a), the inner objective in (5) is strongly concave in z with modulus μ, and H_μ has a $(1/\mu)$-Lipschitz derivative. Nesterov's smoothing theorem (2005, Thm. 1) then applies. For (b), use $(t)_+ - \mu/2 \leq H_\mu(t) \leq (t)_+$ and $\| z \|^2 \leq \sum_i z_i \leq \Gamma$ on Z_Γ. ▫

Fixed smoothing (FS-μ). Run accelerated projected gradient (FISTA with backtracking on L) on φ_μ. By Lemma 1, after k iterations

$$\varphi(x_k) - \varphi^* \leq \mu D + O\left(\frac{L_\mu R^2}{k^2} \right) = \mu D + O\left(\frac{R^2}{\mu k^2} \right).$$

The first term is a floor that does not shrink with k. The choice $\mu \propto 1/k$ balances the two terms and gives the $O(1/k)$ rate of (Nesterov, 2005), but only when the iteration budget is known in advance.

Adaptive smoothing (AS). Use the same iteration with a decreasing schedule $\mu_k = \max\{\mu_0/(1 + k/10), \mu_{\min}\}$, with $\mu_0 = 1$ and $\mu_{\min} = 10^{-3}$, and do not restart the momentum. Early iterations see a heavily smoothed, well-conditioned problem and move fast. Later iterations see a nearly exact one. The schedule needs no iteration budget and suits the rolling setting, where the number of iterations available per epoch varies.

Related methods. Problem (2) could also be attacked as a saddle-point problem with primal–dual splitting (Chambolle & Pock, 2011) or mirror-prox (Nemirovski, 2004). Both need a prox-friendly coupling. Here L is bilinear in $\big(z, h(x)\big)$ but h is non-smooth, so those methods would require an extra lifting of the shortfall variables. Smoothing keeps the problem in the original variables.

5. Experiments

Instance. We use $n = 50$ services and NumPy seed 5. Parameters are drawn as $\bar{d}_i \sim U[5,15]$, $s_i \sim U[0.5,1.5]$ and $\bar{u}_i \sim U[1,3]$. The budget $B = 0.8 \sum_i \bar{d}_i /s_i = 403.43$ covers 80% of full service, and caps are $\bar{x}_i = 2\bar{d}_i/s_i$. The uncertainty parameters are $\rho = 0.5$ and $\Gamma = 10$, so that up to a fifth of the services can be at their worst. The reference value $\varphi^* = 142.7706$ is obtained from LP (4) with the HiGHS solver. All methods start from the projected uniform allocation, where $\varphi = 307.80$.

5.1. Solver comparison

Table 1: Relative gap $(\varphi_{\text{best}} - \varphi^*)/\varphi^*$ after k iterations. SG uses $R = 0.1 \parallel \bar{x} \parallel$. FISTA iterations include backtracking, so each costs at least two oracle calls, against one for SG.

Method	$k = 10$	$k = 50$	$k = 100$	$k = 500$	$k = 1000$	$k = 2000$
SG (Danskin subgradient)	2.1e−1	5.2e−2	2.3e−2	9.1e−3	5.6e−3	2.9e−3
FS, $\mu = 1$	2.0e−1	1.0e−1	9.8e−2	9.8e−2	9.8e−2	9.8e−2
FS, $\mu = 0.1$	2.6e−1	4.7e−2	1.5e−2	1.0e−2	1.0e−2	1.0e−2
FS, $\mu = 0.01$	2.6e−1	1.7e−1	1.1e−1	2.0e−3	1.0e−3	1.0e−3
AS (adaptive μ)	1.8e−1	2.2e−2	1.1e−2	2.0e−3	1.0e−3	5.0e−4

The fixed-smoothing rows show the two terms of the bound in Section 4.2 one at a time. With $\mu = 1$ the method converges quickly to a floor near 10^{-1} and stays there. With $\mu = 0.01$ the floor is two orders of magnitude lower, but reaching it takes several hundred iterations because the problem is badly conditioned ($L_\mu \propto 1/\mu$). The floors scale almost exactly with μ (9.8e−2, 1.0e−2, 1.0e−3), as Lemma 1(b) predicts.

Adaptive smoothing takes the better part of each. At $k = 50$ it is at 2.2e−2, close to the fast $\mu = 0.1$ run. At $k = 1000$ it has caught up with the $\mu = 0.01$ floor, and by $k = 2000$ it is below every fixed choice. Against SG it is better at every checkpoint after $k = 10$, and about 6 times better at $k = 2000$ (5.0e−4 against 2.9e−3). Part of that advantage is offset by the extra oracle calls per iteration noted under Table 1.

5.2. What robustness buys

We compare the nominal allocation ($\Gamma = 0$) with the robust allocation ($\Gamma = 10$) for three uncertainty magnitudes ρ. Each allocation is evaluated in three ways: at the nominal penalties ($z = 0$); at the worst case in Z_{10}; and over 2,000 random scenarios in which 10 services, chosen uniformly, receive $z_i \sim U[0,1]$ (seed 9). We report the mean and the 95th percentile of the random scenarios. “Short” counts the services with positive unmet demand.

Table 2: Nominal against robust allocation, with $\Gamma = 10$ for evaluation.

ρ	Allocation	Nominal cost	Worst-case cost	Random mean	Random 95th pct.	Short services
0.5	Nominal	97.72	146.59	102.60	110.89	9
0.5	Robust	102.32	142.77	107.56	115.24	17
1.0	Nominal	97.72	195.45	107.47	124.06	9
1.0	Robust	117.37	175.18	129.22	138.42	24
2.0	Nominal	97.72	293.17	117.21	150.39	9
2.0	Robust	134.77	228.13	161.86	177.17	32

Three observations follow.

Robustness propagates scarcity. The robust plan arrives 17, 24 and 32 services short, compared to 9 for the nominal plan. This is the type B kink of Section 3 seen in the data. The planner 'morphs' shortfalls to make sure that the adversary does not benefit from concentrating its budget.

The worst-case gain increases with ρ. The decrease in worst case cost is 2.6% at ρ=0.5, 10.4% at ρ=1 and 22.2% at ρ=2. The difference in nominal-cost premium is 4.7%, 20.1% and 37.9%. This is the penalty one pays for robustness as defined by (Bertsimas & Sim, 2004).

Normal scenarios support the nominal plan. On random scenarios (98% of the worst case!) the robust plan costs more, on average and at the 95th percentile, across all ρ. Robust allocation protects against coordinated bad luck, and is worth paying for when that is the real risk, such as a single regional incident affecting many services. In reality, Γ and ρ should be chosen based on the joint distribution of the histories of the penalty shocks rather than on the marginal ranges. You can see the cost of being prepared for a coordination failure that never occurs in the random-scenario columns of Table 2.

Conclusions and limitations

Worst-case cost of an allocation under coupled budgeted uncertainty is the sum of the Γ biggest weighted shortfalls, plus an inconsequential constant. It is convex piecewise linear, and the robust optimum usually lies on a kink in the rank ordering, where the planner has made many shortfalls have the same value to the opponent. This is one way we get distribution of scarcity across items; this is also why it is truly non-smooth, not just "bad at a few points".

The smoothing is useful for computation only if the parameter is required to diminish. Keeping it fixed imposes a lower bound about equal to its value. A straightforward decreasing schedule breaks this lower bound and was roughly six times more accurate than the subgradient method after 2000 iterations, with each iteration incurring additional oracle calls.

The limitations are concrete.

1. One example. For the tester comparison we use one random instance. We use random restarts and larger n to give an indication of stability of the ordering in Table 1.
2. No AS convergence proof. The adaptive schedule is a heuristic in this paper. Its non-restarted momentum is not covered by the fixed-μ bound, and proof would need an analysis of smoothing schemes with non-constant parameters.
3. Single epoch only. While the adaptive (rolling) setting of Section 2 was the motivation for our choice of approaches, we tested for one epoch only. A rolling experiment with forecast updates with warm starts is the natural next step.
4. Linear service curves. The concave service curves si (xi) remain as convex functions on the problem. A non-concave curve would be a weakly convex function, which enables to use the theory of (Davis & Drusvyatskiy, 2019).
5. Data and code availability. All data are artificial, derived from the specific seeds reported (5 for the instance, 9 for the random simulations). Python code to generate Tables 1 and 2 is available from the author upon request.
6. Conflict of interest. The author has stated explicitly that there is no conflict of interest.

References

1. Ben-Tal, A., & Nemirovski, A. (1998). Robust convex optimization. *Mathematics of operations research*, *23*(4), 769-805. https://doi.org/10.1287/moor.23.4.769
2. Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations research*, *52*(1), 35-53. https://doi.org/10.1287/opre.1030.0065
3. Chambolle, A., & Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of mathematical imaging and vision*, *40*(1), 120-145. https://doi.org/10.1007/s10851-010-0251-1
4. Clarke, F. H. (1990). *Optimization and nonsmooth analysis*. Society for industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611971309
5. Danskin, J. M. (1966). The theory of max-min, with applications. *SIAM Journal on Applied Mathematics*, *14*(4), 641-664. https://doi.org/10.1137/0114053
6. Davis, D., & Drusvyatskiy, D. (2019). Stochastic model-based minimization of weakly convex functions. *SIAM Journal on Optimization*, *29*(1), 207-239. https://doi.org/10.1137/18M1178244
7. Kelly, F. P., Maulloo, A. K., & Tan, D. K. H. (1998). Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research society*, *49*(3), 237-252. https://doi.org/10.1057/palgrave.jors.2600523
8. G S Murugapandian. (2026). Applied Nonlinearity in Modeling Secure Internet Service Dynamics for Academic Library and Research Data Systems. *Applied Nonlinearity in Science and Technology*, *2*(2), 1–8.
9. Nesterov, Y. (2005). Smooth minimization of non-smooth functions. *Mathematical programming*, *103*(1), 127-152. https://doi.org/10.1007/s10107-004-0552-5
10. Palomar, D. P., & Chiang, M. (2006). A tutorial on decomposition methods for network utility maximization. *IEEE Journal on Selected Areas in Communications*, *24*(8), 1439-1451. https://doi.org/10.1109/JSAC.2006.879350
11. Goswami, A. (2026). Certain ways of constructing new spaces using∆-approach. *Results in Nonlinear Analysis*, *9*(2), 179-185.
12. Dr. K. Chitra. (2026). Oscillatory Failure Prediction in Lithium-Ion Battery Systems Through Nonlinear State Coupling. *Applied Nonlinearity in Science and Technology*, *2*(2), 29–34.
13. Deepika, J. (2026). Nonlinear Mathematical Model for Dynamic Frequency Stability Analysis of Renewable-Dominated Power Systems. *Journal of Applied Mathematical Models in Engineering*, 10-20.
14. Dr. Idris Mansour. (2026). Dynamic Modeling and Intelligent Control of an Autonomous Mobile Manipulator for Precision Industrial Material Handling. *Advances in Mechanical Engineering and Applications*, *2*(3), 41-47.
15. H. K. Mzeh, L. Salabi, M. T. Jafer, Sikalu T C. (2026). Intelligent Assistive Communication Models for Personalized Learning Environments. *Journal of Intelligent Assistive Communication Technologies*, *2*(1), 64-71.