Поиск равновесий в многостадийных транспортных моделях

А.В. Гасников (ПреМоЛаб МФТИ, ИППИ РАН, ВШЭ, ИПМ РАН)

П.Е. Двуреченский (ИППИ РАН, ПреМоЛаб МФТИ, WIAS Berlin)

П.И. Стецюк (Институт кибернетики, НАН Украины)

А.Л. Суворикова (ПреМоЛаб МФТИ, IRTG 1792)

А.В. Чернов (Кафедра МОУ ФУПМ МФТИ)

Поиск (стохастических) равновесий в многостадийных моделях транспортных потков приводит к решению следующей седловой задачи с правильной (выпукло-вогнутой) структурой [1, 2]:

$$\min_{\substack{\sum_{j=1}^{n} x_{ij}=L_i,\ \sum_{i=1}^{n} x_{ij}=W_j \\ i,j=1,\dots,n}} \max_{y\in Q} \left\{ \sum_{i,j=1}^{n} x_{ij} \ln x_{ij} + \sum_{i,j=1}^{n} c_{ij}(y)\, x_{ij} + g(y) \right\} =$$

$$= \max_{y\in Q} \max_{\lambda,\mu} \left\{ \langle \lambda, L\rangle + \langle \mu, W\rangle - \sum_{i,j=1}^{n} \exp\left(-c_{ij}(y) - 1 + \lambda_i + \mu_j\right) + g(y) \right\} = \max_{y\in Q} f(y),$$

где $g(y)$ и $c_{ij}(y)$ – вогнутые гладкие функции (если ищутся не стохастические равновесия, то $c_{ij}(y)$ могут быть негладкими). В данной работе мы покажем, как можно решать такую задачу.

Замечание 1. К похожей задаче приводит поиск барицентра Монжа–Канторовича (в западной литературе чаще говорят барицентра Вассерштейна [4]). Для наглядности мы ограничимся здесь рассмотрением частного случая этой задачи – поиска барицентра одной дискретной вероятностной меры W (в этом случае барицентр совпадает с самой мерой, хотя все получается тривиально, но сама схема рассуждений нам представляется полезной, на базе нее строится решение и общей задачи [4 – 6]). Итак, нужно решать задачу

$$\min_{L\in S_n(1)} \max_{\lambda,\mu} \left\{ \langle \lambda, L\rangle + \langle \mu, W\rangle - \sum_{i,j=1}^{n} \exp\left(-c_{ij} - 1 + \lambda_i + \mu_j\right) \right\} =$$

$$= \min_{L\in S_n(1)} \max_{\lambda} \left\{ \langle \lambda, L\rangle - \underbrace{\sum_{j=1}^{n} W_j \ln\left(\frac{1}{W_j} \sum_{i=1}^{n} \exp\left(-c_{ij}+\lambda_i\right)\right)}_{\varphi^*(\lambda)} \right\} = \min_{L\in S_n(1)} \varphi(L) =$$

$$= -\max_{L\in S_n(1)} \left\{ \langle 0, L\rangle - \varphi(L)\right\} = -\varphi^*(0), \quad L_* = \arg\min_{L\in S_n(1)} \varphi(L) = \nabla\varphi^*(0),$$

где $L, W \in S_n(1)$ – единичный симплекс. Хотя явно вычислить функцию $\varphi(L)$ мы не можем, зато у нас есть явная формула для сопряженной функции $\varphi^*(\lambda)$ и ее градиента. Следовательно, в данном случае задачу можно решить явно. Ответ получается ожидаемым $L = W$. В общем случае, как правило, явно решить такого рода седловую задачу не удается. Теперь мы переходим к изложению общего случая.

Из замечания 1 следует, что внутренняя задача максимизации по (λ, μ) может быть явно решена по μ при фиксированном λ, и наоборот. Собственно, таким образом, получается метод балансировки расчета матрицы корреспонденций по энтропийной модели [3] (тесно связанный с методом Синхорна [5, 7]), как метод простой итерации для явно выписываемых условий экстремума (принципа Ферма): $\lambda = \Lambda(\mu)$, $\mu = \mathrm{M}(\lambda)$. Оператор $(\lambda, \mu) \to (\Lambda(\mu), \mathrm{M}(\lambda))$ является сжимающим в метрике Биркгофа–Гильберта ρ [8]. Это означает, после $N \sim \ln(\sigma^{-1})$ итераций метода балансировки можно получить такие (λ_N, μ_N), что $\rho((\lambda_N, \mu_N); (\lambda_*(y), \mu_*(y))) \le \sigma$. Причем на практике наблюдается очень быстрая сходимость, то есть коэффициент пропорциональности не большой [3]. Таким образом, мы можем приближенно решить внутреннюю задачу. Далее предлагается воспользоваться универсальным методом Ю.Е. Нестерова [9] для решения внешней задачи максимизации по y. К сожалению, кроме того что внешняя задач гладкая, больше мы ничего о ней сказать не можем. В частности, не знаем константу Липшица градиента функционала внешней задачи, не знаем размер решения. Поэтому и отдано предпочтение универсальному методу. Однако нам потребуется использовать этот метод в варианте с неточным оракулом, выдающим градиент [10].

Определение [11]. (δ, L)*-оракул выдает (на запрос, в котором указывается только одна точка* y*) такие* $(F(y), G(y))$*, что и для любых* $y, y' \in Q$

$$0 \le f(y') - F(y) - \langle G(y), y' - y \rangle \le \frac{L}{2}\|y' - y\|^2 + \delta.$$

Заметим [9], что можно погрузить задачу с гельдеровым градиентом ($\nu \in [0,1]$)

$$\|\nabla f(y') - \nabla f(y)\|_* \le L_\nu \|y' - y\|^\nu$$

(в том числе и негладкую задачу с ограниченной нормой разности субградиентов при $\nu = 0$) в класс гладких задач с оракулом, характеризующимся точностью δ и

$$L = L_\nu \left[\frac{L_\nu (1-\nu)}{2\delta(1+\nu)} \right]^{\frac{1-\nu}{1+\nu}}.$$

Итак, у нас есть внешняя задача $\max\limits_{y \in Q} f(y)$, для которой обращение к (δ, L) оракулу за значением функции и градиента стоит $\sim \ln(\delta^{-1})$. Насколько быстро мы можем решить такую задачу, т.е. при каком $N(\varepsilon)$ можно гарантировать, что

$$\max_{y \in Q} f(y) - f\left(y_{N(\varepsilon)}\right) \le \varepsilon ?$$

Ответ можно получить из следующего результата.

Теорема (см. [10]). *Существует однопараметрическое семейство универсальных градиентных методов (параметр $p \in [0,1]$), не получающих на вход кроме p больше никаких параметров (в частности, не использующих значения L_ν и R – "расстояние" от точки старта до решения, априорно не известное), которое приводит к следующей оценке на требуемое число итераций*

$$N_p(\varepsilon) = \mathrm{O}\left(\inf_{\nu \in [0,1]} \left(\frac{L_\nu R^{1+\nu}}{\varepsilon} \right)^{\frac{2}{1+2p\nu+\nu}} \right),$$

если $\delta \le \mathrm{O}\left(\varepsilon / N_p(\varepsilon)^p\right)$.

Из теоремы можно заключить, что стоит выбирать значение параметра $p = 1$, при этом общее трудозатраты машинного времени будут $\mathrm{O}\left(N_1(\varepsilon) T \ln(\varepsilon^{-1})\right)$, где T – время решения вспомогательной задачи методом балансировки с относительной точностью 1%.

Численные эксперименты показывают, что на одном современном ноутбуке при $n \sim 10^2$ время $T \approx 1$ сек. [10].

Замечание 2. В приложениях возникают задачи, имеющие следующий вид (см., например, [12 – 14]) $f(x) = \Phi(x, y(x)) \to \min\limits_x$ при этом $y(x)$ и $\nabla y(x)$ могут быть получены из решения отдельной подзадачи лишь приближенно. Довольно типично, что существует способ, который выдает ε-приближенное решение за время, зависящее от ε, логарифмическим образом $\sim \ln(\varepsilon^{-1})$. В данной работе намечен общий способ решения таких задач. Его наиболее важной отличительной чертой является адаптивность (самонастраиваемость), т.е. методу на вход не надо подавать никаких констант Липшица (более того метод будет работать и в негладком случае). Метод сам настраивается локально на оптимальную гладкость функции. Эти свойство метода делает его привилегированным, поскольку в реальных приложениях, чтобы что-то знать о свойствах $f(x)$ нужно что-то знать о свойствах зависимости $y(x)$, что часто не доступно по постановке задачи, или, при попытке оценить, приводит к сильно завышенным оценкам.

Замечание 3. В задаче из замечания 2 важно уметь эффективно пересчитывать значения $y(x)$, а не рассчитывать их каждый раз заново (на каждой итерации внешнего цикла). Поясним сказанное. Предположим, что мы уже как-то посчитали, скажем, $y(x)$, решив, например, с какой-то точностью соответствующую задачу оптимизации. Тогда для вычисления $y(x+\Delta x)$ (на следующей итерации внешнего цикла) у нас будет хорошее начальное приближение $y(x)$. А, как известно (см., например, [13]), расстояние от точки старта до решения (не в случае сходимости метода со скоростью геометрической прогрессии или быстрей) существенным образом определяет время работы алгоритма. Тем не менее, известные нам приложения (см., [12 – 14]) пока как раз всецело соответствуют сходимости процедуры поиска $y(x)$ со скоростью геометрической прогрессии. Связано это с тем, что если расчет $y(x)$ с точностью ε осуществляется за $O(\ln(\varepsilon^{-1}))$ операций, то для внешней задачи можно выбирать самый быстрый метод (а, стало быть, и самый требовательный к точности), и с точностью до того, что стоит под логарифмом, общая трудоемкость будет просто прямым произведением трудоемкостей решения внутренней и внешних задач по отдельности. Как правило, такое сочетание оказывается недоминируемым.

Замечание 4. В реальных транспортных приложениях [1, 2] наиболее сложным является расчет $c_{ij}(y)$ и их градиентов (особенно при поиске стохастических равновесий). Тем не менее, эти задачи имеют вполне четкую привязку к решению некоторых задач на графах типа поиска кратчайших путей. Также как и в предыдущем замечании для внутренней задачи, для внешней задачи, можно не рассчитывать $c_{ij}(y)$ и их градиенты каждый раз заново, а пересчитывать; также можно допускать неточность в их вычислении (и ненулевую вероятность ошибки) надеясь на ускорение (благо метод работает в концепции неточного оракула, природа которой не принципиальна [11]). Также здесь оказываются полезными идеи БАД (быстрого автоматического дифференцирования [15, 16]), которые позволяют практически за тоже время, что занимает вычисление самих функций, вычислять их градиенты.

Авторы выражают благодарность Ю.Е. Нестерову за ряд ценных замечаний.

Исследование А.В. Гасникова, П.Е. Двуреченского были выполнены в ИППИ РАН за счет гранта Российского научного фонда (проект № 14-50-00150), и частично выполнено в МФТИ при поддержке грантов РФФИ 15-31-20571-мол_а_вед, 14-01-00722-а, 13-01-12007-офи_м.

Литература

1. *Гасников А.В., Дорн Ю.В., Нестеров Ю.Е, Шпирко С.В.* О трехстадийной версии модели стационарной динамики транспортных потоков // Математическое моделирование. 2014. Т. 26:6. С. 34–70. arXiv:1405.7630
2. *Гасников А.В.* Об эффективной вычислимости конкурентных равновесий в транспортно-экономических моделях // Математическое моделирование. 2015. Т. 27. (в печати) arXiv:1410.3123
3. *Гасников А.В., Гасникова Е.В., Нестеров Ю.Е., Чернов А.В.* Об эффективных численных методах решения задач энтропийно-линейного программирования // ЖВМ и МФ. 2016. (принята к печати) arXiv:1410.7719
4. *Cuturi M., Doucet A.* Fast Computation of Wasserstein Barycenters // ICML, 2014. http://www.iip.ist.i.kyoto-u.ac.jp/member/cuturi/
5. *Benamou J.D., Carlier G., Cuturi M., Nenna L., Peyré G.* Iterative Bregman Projections for Regularized Transportation Problems // e-print, 2015. arXiv:1412.5154 (to appear in SISC)
6. *Cuturi M., Peyré G., Rolet A.* A Smoothed Dual Formulation for Variational Wasserstein Problems // e-print, 2015. arXiv:1503.02533
7. *Cuturi M.* Sinkhorn Distances: Lightspeed Computation of Optimal Transport // NIPS, 2013.
8. *Franklin J., Lorenz J.* On the scaling of multidimensional matrices // Linear Algebra and its applications. 1989. V. 114. P. 717–735.

9. *Nesterov Yu.* Universal gradient methods for convex optimization problems // CORE Discussion Paper 2013/63. 2013.
10. *Гасников А.В, Двуреченский П.Е., Камзолов Д.И.* Градиентные и прямые методы с неточным оракулом для задач стохастической оптимизации // Труды SDCP-2014. Екатеринбург, 2015. arXiv:1502.06259
11. *Devolder O.* Exactness, inexactness and stochasticity in first-order methods for large-scale convex optimization. CORE UCL, PhD thesis, March 2013. http://www.ecore.be/DPs/dp_1327057920.pdf
12. *Dvurechensky P., Gasnikov A., Zhukovskii M.* Random gradient-free methods for random walk based web page ranking functions learning // e-print, 2014. arxiv:1411.4282
13. *Гасников А.В., Двуреченский П.Е., Нестеров Ю.Е.* Стохастические градиентные методы с неточным оракулом // Автоматика и телемеханика. 2016. (принята к печати) arxiv:1411.4218
14. *Гасников А.В., Гасникова Е.В., Двуреченский П.Е.* Современные подходы к численному поиску стохастического равновесия в модели равновесного распределения транспортных потоков Бэкмана // ЖВМ и МФ. 2016. (подана) arXiv:1505.07492
15. *Ким К., Нестеров Ю., Скоков В., Черкасский Б.* Эффективные алгоритмы для дифференцирования и задачи экстремали // Экономика и математические методы. 1984. Т. 20. С. 309–318.
16. *Евтушенко Ю.Г.* Оптимизация и быстрое автоматическое дифференцирование. М.: ВЦ РАН, 2013.