math.e

Hrvatski matematički elektronički časopis

Uparena optimizacijska metoda

gradijentni i zrcalni spust **hibridna ili uparena metoda** **konveksna optimizacija**

Luka Borozan, Slobodan Jelić, Domagoj Matijević, Domagoj Ševerdija
Sveučilište J.J. Strossmayera u Osijeku, Odjel za matematiku
lborozan@mathos.hr

Sažetak

U ovom članku analiziramo metode gradijentnog i zrcalnog spusta u području konveksne optimizacije s danim naglaskom na njihove brzine konvergencije. Nadalje, uparujući dvije spomenute metode dobivamo takozvanu uparenu metodu čija analiza konvergencije pokazuje ubrzanje u odnosu na gradijentnu i zrcalnu metodu, te bilo koju drugu nama poznatu metodu prvoga reda.

Uvod

Probleme konveksne optimizacije često nalazimo kako u različitim područjima matematike i računarstva, tako i u primjeni. Do sada su razvijene mnoge metode konveksne optimizacije, primjerice metoda gradijentnog spusta, metoda najbržeg spusta, metoda zrcalnog spusta, metoda polovljenja, itd. Više o tim metodama može se naći u [4], [3]. Posebno ćemo promatrati takozvane metode prvoga reda koje se odlikuju jednostavnošću implementacije na računalu. U ovom ćemo radu pokazati kako uparivanjem metoda gradijentnog i zrcalnog spusta dobivamo takozvanu uparenu metodu koja brže konvergira prema rješenju od bilo koje nama poznate metode prvog reda.

Prvo ćemo ukratko objasniti metode gradijentnog i zrcalnog spusta, a potom ćemo konstruirati i opravdati uparenu metodu.

Metoda gradijentnog spusta

Metoda gradijentnog spusta ili gradijentna metoda jedna je od najstarijih i najjednostavnijih iterativnih metoda rješavanja problema konveksne optimizacije. Ona minimizira konveksnu, diferencijabilnu funkciju f po varijabli $x \in S$, gdje je S konveksan i zatvoren skup. Dodatno zahtijevamo da je ∇f Lipschitz neprekidna na S s konstantom L. Prvo ćemo dati definiciju koraka (iteracije) metode gradijentnog spusta.

Definicija 1. *Neka je $x \in S$. **Iteracija gradijentne metode** s duljinom koraka $1/L$ je*

$$\tilde{x} = \mathrm{Grad}(x) = \mathrm{argmin}_{y \in S} \left\{ f(x) + \frac{L}{2} \|y - x\|^2 + \langle \nabla f(x), y - x \rangle \right\}.$$

Označimo $\mathrm{Prog}(x) = -\min_{y \in S} \left\{ \frac{L}{2} \|y - x\|^2 + \langle \nabla f(x), y - x \rangle \right\} \geq 0$. Primijetimo kako se gornja definicija koraka gradijentne metode bitno razlikuje od gradijentne metode kod koje je $S = \mathbb{R}^n$ i $\|\cdot\| = \|\cdot\|_2$, tj. kod gradijentne metode bez ograničenja u euklidskoj normi ($\mathrm{Grad}(x) = x - \frac{1}{L} f'(x)$). Više o gradijentnoj metodi u neeuklidskim normama može se pronaći u [5]. Naime, različite norme rezultiraju različitim gradijentnim metodama. Trebat će nam sljedeća definicija.

Definicija 2. *Neka je $x \in \mathbb{R}^n$, te $\|\cdot\|$ norma na $\mathbb{R}^n$. **Dualnu normu** norme $\|\cdot\|$ definiramo kao*

$$\|x\|_* = \max_{\|v\|=1} v^T x.$$

Sljedeći teorem daje nam ocjenu smanjenja vrijednosti funkcije cilja metode gradijentnog spusta, te je dokazan po uzoru na [2].

Teorem 3. *Neka je $x \in S$. Vrijedi $f(\mathrm{Grad}(x)) \leq f(x) - \mathrm{Prog}(x)$.*

Dokaz. Neka je $\tilde{x} = \mathrm{Grad}(x)$. Imamo:

$$\begin{aligned}
\mathrm{Prog}(x) &= -\min_{y \in S}\left\{\frac{L}{2}\|y-x\|^2 + \langle \nabla f(x), y-x\rangle\right\} \\
&= -\left(\frac{L}{2}\|\tilde{x}-x\|^2 + \langle \nabla f(x), \tilde{x}-x\rangle\right) \\
&= f(x) - \left(\frac{L}{2}\|\tilde{x}-x\|^2 + \langle \nabla f(x), \tilde{x}-x\rangle + f(x)\right) \leq f(x) - f(\tilde{x}).
\end{aligned}$$

Posljednja nejednakost posljedica je sljedećeg izraza. Neka su $x, y \in S$.

$$\begin{aligned}
f(y) - f(x) &= \int_0^1 \langle \nabla f(x + t(y-x)), y-x\rangle \mathrm{d}t \\
&= \langle \nabla f(x), y-x\rangle + \int_0^1 \langle \nabla f(x+t(y-x)) - \nabla f(x), y-x\rangle \mathrm{d}t \\
&\leq \langle \nabla f(x), y-x\rangle + \int_0^1 \|\nabla f(x+t(y-x)) - \nabla f(x)\|_* \|y-x\| \,\mathrm{d}t \\
&\leq \langle \nabla f(x), y-x\rangle + \int_0^1 tL\|y-x\|^2 \mathrm{d}t \\
&= \langle \nabla f(x), y-x\rangle + \frac{L}{2}\|y-x\|^2.
\end{aligned}$$

Time je tvrdnja dokazana. ∎

Sada ćemo dati i ocjenu pogreške gradijentne metode.

Teorem 4. *([2]) Neka je x_k aproksimacija točke minimuma x^* dobivena k-tom iteracijom metode gradijentnog spusta s početnom aproksimacijom x_0. Tada vrijedi $f(x_k) - f(x^*) \leq \mathcal{O}\left(\frac{L\|x_0 - x^*\|_2^2}{k}\right)$. Nadalje, ε-aproksimaciju točke x^* postižemo u $\Omega(\frac{L\|x_0-x^*\|_2^2}{\varepsilon})$ iteracija.*

Metoda zrcalnog spusta

Kao i kod metode gradijentnog spusta, tako i kod metode zrcalnog spusta pretpostavljamo da je f konveksna funkcija u varijabli $x \in S$, gdje je S konveksan i zatvoren skup, te da je Lipschitz neprekidna na S s konstantom K. Kako bismo mogli uvesti ovu metodu, potrebno nam je definirati nekoliko pojmova.

Definicija 5. *Neka je f konveksna funkcija na $\mathbb{R}^n$. Kažemo da je $\partial f(x)$ **subgradijent** od f u x ako je $f(y) \geq f(x) + \langle \partial f(x), y - x \rangle, \quad \forall y \in \mathbb{R}^n$.*

Primjer 6. *Za funkciju $f(x) = |x|$ subgradijent ∂f možemo definirati na sljedeći način. Neka je $x \in \mathbb{R}$. Ako je $x < 0$, očito je $\partial f(x) = -1$, a ako je $x > 0$, očito je $\partial f(x) = 1$. Preostaje odrediti $\partial f(0)$. Mora vrijediti $|x| \geq \partial f(0) \cdot x$. To je zadovoljeno za $\partial f(0) \in \{-1, 1\}$.*

Sljedeća propozicija govori o subgradijentnom uvjetu ekvivalentnom Lipschitzovoj neprekidnosti kojeg ćemo dalje koristiti kako bismo definirali korak metode zrcalnog spusta.

Propozicija *Neka je f konveksna funkcija na konveksnom skupu S. Tada je f Lipschitz neprekidna s konstantom K na S ako i samo ako za svaku točku $x \in S$ vrijedi $\|\partial f(x)\|_* \leq K$, gdje je ∂f bilo koji subgradijent od f.*

$(\Rightarrow)$ Pretpostavimo da je f Lipschitz neprekidna s konstantom K na S, te $\{x, y \in S\}$. Neka je $y - x = \operatorname{argmax}_{\|z\|=1} \langle \partial f(x), z \rangle$. Tada vrijedi $\|y - x\| = 1$, te $\langle \partial f(x), z \rangle = \|\partial f(x)\|_*$ po definiciji dualne norme. Koristeći pretpostavku o Lipschitz neprekidnosti od f, definiciju subgradijenta, te Cauchy-Schwarz-Bunyakovsky nejednakost imamo:

$$K\|y-x\| \geq f(y)-f(x) \geq \langle \partial f(x), y-x\rangle = \|\partial f(x)\|_*.$$

Iz $\|y-x\|=1$ *slijedi prva implikacija.*

$(\Leftarrow)$ *Neka sada* $\forall x \in S$ *vrijedi* $\|\partial f(x)\|_* \leq K$. *Neka je* $y \in S$. *Množeći definiciju subgradijenta s* -1, *te primjenom Cauchy-Schwarz-Bunyakovsky nejednakosti imamo:*

$$f(x)-f(y) \leq \langle \partial f(x), x-y\rangle \leq \|\partial f(x)\|_* \|x-y\| \leq K\|x-y\|.$$

Analogno, zamjenom uloga x *i* y *dobivamo* $f(y)-f(x) \leq K\|x-y\|$. *Uparivanjem prethodne dvije nejednakosti dobivamo:*

$$\max\{f(y)-f(x), f(x)-f(y)\} = |f(x)-f(y)| \leq K\|x-y\|.$$

Time smo dokazali i drugu implikaciju, pa tako i tvrdnju propozicije. ■

U svrhu definiranja iteracije metode zrcalnog spusta trebaju nam i sljedeće definicije.

Definicija 7. *Kažemo da je* $d : \mathbb{R}^n \times \mathbb{R}^n \to [0,\infty\rangle$ ***kvazimetrička funkcija*** *ukoliko* $(\forall x, y \in \mathbb{R}^n)$ *vrijedi* $d(x,y)=0 \iff x=y$.

Definicija 8. *Neka je* ϕ *1-jako konveksna, diferencijabilna funkcija na* $\mathrm{Int}S$. *Funkciju* $d : \mathrm{Int}S \times S \to [0,\infty\rangle$ *definiranu s* $d(x,y)=\phi(y)-\phi(x)-\langle \nabla\phi(x), y-x\rangle$ *nazivamo* ***Bregmanova divergencija***.

Napomena Vrijedi $d(x,y) \geq \dfrac{1}{2}\|x-y\|^2$. Dokaz ove tvrdnje ćemo izostaviti.

Sada možemo definirati korak metode zrcalnog spusta.

Definicija 9. ***Iteracija metode zrcalnog spusta*** *s duljinom* $\alpha > 0$ *dana je sljedećim izrazom:*

$$\mathrm{Mirr}(\alpha, x, \partial f) = \mathrm{argmin}_{y\in S}\{d(x,y) + \langle \alpha\partial f(x), y-x\rangle\}$$

gdje je ∂f *neki subgradijent funkcije* f.

Sljedeći rezultat govori o brzini konvergencije metode zrcalnog spusta. Dokaz mu je preuzet iz [2].

Teorem 10. *Ako je $\tilde{x} = \mathrm{Mirr}(\alpha, x, \partial f)$, gdje su $\alpha > 0$ duljina koraka, te ∂f neki subgradijent od f, tada $\forall u \in S$ vrijedi:*

$$\alpha(f(x) - f(u)) \leq \alpha\langle \partial f(x), x - u\rangle \leq \frac{\alpha^2}{2}\|\partial f(x)\|_*^2 + d(x,u) - d(\tilde{x}, u).$$

Dokaz. Prva nejednakost slijedi iz definicije subgradijenta. Počnimo od izraza:

$$\langle \alpha\partial f(x), x - u\rangle \quad = \quad \langle \alpha\partial f(x), x - \tilde{x}\rangle + \langle \alpha\partial f(x), \tilde{x} - u\rangle.$$

Iz činjenice da je $\tilde{x}$ dobiven minimizacijom po $y \in S$ izraza $d(x,y) + \langle \alpha\partial f(x), y\rangle$, dobivamo da je $\langle \partial_2 d(x,\tilde{x}) + \alpha\partial f(x), u - \tilde{x}\rangle = 0$, pa imamo:

$$\langle \alpha\partial f(x), x - u\rangle \quad \leq \quad \langle \alpha\partial f(x), x - \tilde{x}\rangle + \langle -\partial_2 d(x,\tilde{x}), \tilde{x} - u\rangle.$$

Ocijenimo sada drugi sumand na lijevoj strani prethodne nejednakosti.

$$\begin{aligned}
\langle -\partial_2 d(x,\tilde{x}), \tilde{x} - u\rangle &= \langle \nabla\phi(x) - \nabla\phi(\tilde{x}), \tilde{x} - u\rangle \\
&= (\phi(u) - \phi(x) - \langle \nabla\phi(x), u - x\rangle) - (\phi(u) - \phi(\tilde{x}) - \langle \nabla\phi(\tilde{x}), u - \tilde{x}\rangle) \\
&\quad -(\phi(\tilde{x}) - \phi(x) - \langle \nabla\phi(x), \tilde{x} - x\rangle) \\
&= d(x,u) - d(\tilde{x},u) - d(x,\tilde{x}) \leq d(x,u) - d(\tilde{x},u) - \frac{1}{2}\|x - \tilde{x}\|^2.
\end{aligned}$$

Posljednja nejednakost slijedi iz svojstva Bregmanove divergencije danog u 8. Sada imamo:

$$\begin{aligned}
\langle \alpha\partial f(x), x - u\rangle &\leq (\langle \alpha\partial f(x), x - \tilde{x}\rangle - \frac{1}{2}\|x - \tilde{x}\|^2) + (d(x,u) - d(\tilde{x},u)). \\
&\leq \frac{\alpha^2}{2}\|\partial f(\tilde{x})\|_*^2 + (d(x,u) - d(\tilde{x},u)).
\end{aligned}$$

Posljednja nejednakost slijedi iz $ab - \frac{1}{2}b^2 \leq \frac{1}{2}a^2,\ a, b \in \mathbb{R}$. Time je teorem dokazan. ■

Za kraj, navedimo rezultat koji ocjenjuje pogrešku metode zrcalnog spusta.

Teorem 11. *Neka je $\bar{x} = \frac{1}{k}\sum_{i=0}^{k-1} x_i$ aritmetička sredina prvih k iteracija metode zrcalnog spusta, Θ bilo koja gornja međa na $d(x_0, x^*)$, gdje je x^* točka minimuma, te neka je $\alpha = \frac{\sqrt{2\Theta}}{K\sqrt{k}}$. Vrijedi: $f(\bar{x}) - f(x^*) \leq \frac{\sqrt{2\Theta}K}{\sqrt{k}}$. Nadalje, ε-aproksimaciju točke x^* postižemo u $\Omega(\frac{2\Theta K^2}{\varepsilon^2})$ iteracija.*

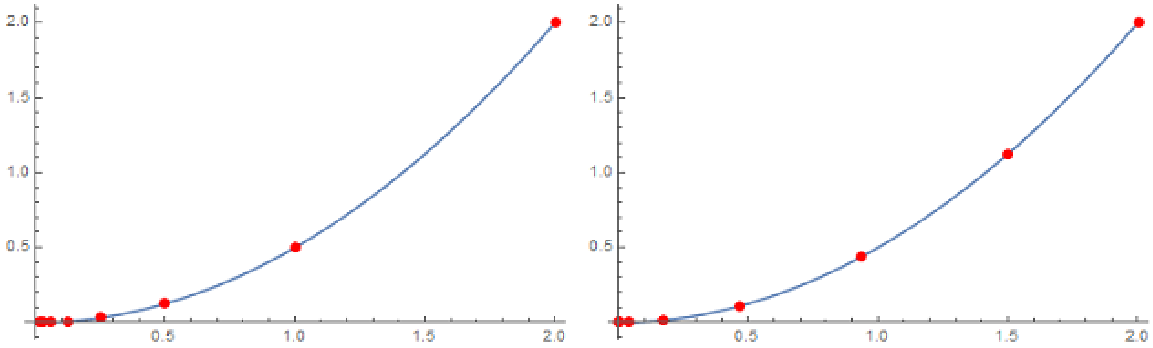
Dokaz prethodnog teorema može se pronaći u [2], dok se više o metodi može pronaći u [3].

Osvrt na metode gradijentnog i zrcalnog spusta

U praksi se pokazalo kako gradijentna metoda brzo konvergira dok je trenutna iteracija ”daleko” od optimalnog rješenja, dok metoda zrcalnog spusta konvergira brzo kada je ”blizu” optimuma. Zbog toga ćemo u nastavku razviti metodu koja će izmjenjivati korake gradijentne i metode zrcalnog spusta ovisno o tome koliko koja metoda pridonosi konvergenciji prema optimumu.

Napomena. U slučaju kada je $\phi(x) = \frac{1}{2}\|x\|_2^2$ i $\|\cdot\| = \|\cdot\|_2$ gradijentne i zrcalne iteracije su ekvivalentne.

Primjer 12. *Metodama gradijentnog i zrcalnog spusta minimizirat ćemo funkciju* $f(x) = \frac{x^2}{2}$ *s početnom aproksimacijom* $x_0 = 2$ *na segmentu* $[0, 2]$*. Istaknute točke na sljedeća dva grafa prikazuju iteracije dviju metoda s desna na lijevo.*



Slika 1: Iteracije metode gradijentnog spusta (lijevo) i zrcalnog spusta (desno)

Na slici 1 se jasno vidi kako je metoda gradijentnog spusta brža od metode zrcalnog spusta u prvih nekoliko iteracija. Također, uočavamo kako metoda zrcalnog spusta brže konvergira nakon prvih nekoliko iteracija. Ovim smo primjerom intuitivno opravdali konstrukciju metode koja uparuje gradijentnu i zrcalnu metodu.

Pojednostavljena uparena metoda

Pokušat ćemo konstruirati metodu koja uparuje metode gradijentnog i zrcalnog spusta s ciljem ubrzanja konvergencije. Ova metoda je prvi puta spomenuta u [1], odakle su preuzeti dokazi većine teorema u ovom i sljedećem potpoglavlju. Napominjemo kako je slična metoda (dualnog usrednjavanja) dana u [7]. Pretpostavke ćemo naslijediti iz metode gradijentnog spusta, tj. pretpostavit ćemo kako je f konveksna, diferencijabilna i Lipschitz neprekidna s konstantom L. Finalna metoda će koristiti varijabilne duljine koraka u svakoj iteraciji, no prije nego to postignemo, konstruirat ćemo jednostavniju metodu s konstantnom duljinom koraka i bez ograničenja ($S = \mathbb{R}^n$).

Ideja pojednostavljenog algoritma je sljedeća:

- postavimo tri početne iteracije $x_0 = y_0 = z_0$,
- u svakoj iteraciji k računamo vrijednosti $x_{k+1}, y_{k+1}, z_{k+1}$,

- $x_{k+1} = \tau z_k + (1-\tau) y_k$,
- $y_{k+1} = \mathrm{Grad}(x_{k+1})$,
- $z_{k+1} = \mathrm{Mirr}(\alpha, x_{k+1}, \partial f)$.

Još nam preostaje razmotriti brojeve α i τ. Broj α je zasada konstantna duljina koraka metode zrcalnog spusta, dok je τ također privremeno konstantna težina pomoću koje kontroliramo utjecaj metoda gradijentnog i zrcalnog spusta na aproksimaciju rješenja. Sljedeći teorem nam daje osnovni rezultat koji će nam pomoći pri dokazivanju teorema o brzini konvergencije pojednostavljene uparene metode.

Teorem 13. *Za svaki u iz $\mathbb{R}^n$ imamo:*

$$\begin{aligned}\alpha\langle\nabla f(x_{k+1}), z_k - u\rangle &\leq \frac{\alpha^2}{2}\|\nabla f(x_{k+1})\|_*^2 + d(z_k, u) - d(z_{k+1}, u)\\ &\leq \alpha^2 L(f(x_{k+1}) - f(y_{k+1})) + d(z_k, u) - d(z_{k+1}, u).\end{aligned}$$

Dokaz. Dokaz prve nejednakosti identičan je dokazu Teorema 10 uz promijenjene oznake. Iz Teorema 3 imamo (uz promjenu oznaka) $f(x_{k+1}) - f(y_{k+1}) \geq \mathrm{Prog}(x_{k+1})$, gdje je

$$\begin{aligned}\mathrm{Prog}(x_{k+1}) &= -\min_{y\in\mathbb{R}^n}\left\{\frac{L}{2}\|y - x_{k+1}\|^2 + \langle\nabla f(x_{k+1}), y - x_{k+1}\rangle\right\}\\ &= -L\min_{y\in\mathbb{R}^n}\left\{\frac{1}{2}\|x_{k+1} - y\|^2 - \frac{1}{L}\langle\nabla f(x_{k+1}), x_{k+1} - y\rangle\right\}\\ &\geq -L\min_{y\in\mathbb{R}^n}\left\{\frac{1}{2}\|x_{k+1} - y\|^2 - \frac{1}{L}\|\nabla f(x_{k+1})\|_*\|x_{k+1} - y\|\right\}\\ &\geq -L\min_{y\in\mathbb{R}^n}\left\{-\frac{1}{2L^2}\|\nabla f(x_{k+1})\|_*^2\right\} = \frac{1}{2L}\|\nabla f(x_{k+1})\|_*^2.\end{aligned}$$

Koristeći nejednakost $f(x_{k+1}) - f(y_{k+1}) \geq \frac{1}{2L}\|\nabla f(x_{k+1})\|_*^2$ slijedi i druga nejednakost. ■

Iz prethodnog teorema vidimo kako je pogreška iteracije metode zrcalnog spusta $\frac{\alpha^2}{2}\|\nabla f(x_{k+1})\|_*^2$ proporcionalna smanjenju vrijednosti funkcije cilja dobivenom gradijentnom iteracijom $f(x_{k+1}) - f(y_{k+1})$. Na taj način opravdavamo praktični rezultat (o brzini konvergencije gradijentne i zrcalne metode) iz prethodnog potpoglavlja.

Sada ćemo navesti rezultat koji daje brzinu konvergencije pojednostavljene uparene metode.

Teorem 14. *Ukoliko odaberemo $\tau \in (0,1)$ takav da je $\frac{1-\tau}{\tau} = \alpha L$, tada $\forall u \in \mathbb{R}^n$ imamo:*

$$\alpha\langle\nabla f(x_{k+1}), x_{k+1} - u\rangle \leq \alpha^2 L(f(y_k) - f(y_{k+1})) + (d(z_k, u) - d(z_{k+1}, u)).$$

Dokaz. Primijetimo da za x_{k+1} vrijedi $\tau(x_{k+1} - z_k) = (1-\tau)(y_k - x_{k+1})$. Koristeći tu činjenicu i osnovna svojstva konveksnih funkcija imamo:

$$\begin{aligned} \alpha\langle\nabla f(x_{k+1}), x_{k+1} - u\rangle - \alpha\langle\nabla f(x_{k+1}), z_k - u\rangle &= \alpha\langle\nabla f(x_{k+1}), x_{k+1} - z_k\rangle \\ &= \frac{1-\tau}{\tau}\alpha\langle\nabla f(x_{k+1}), x_{k+1} - z_k\rangle \\ &\leq \frac{1-\tau}{\tau}\alpha(f(y_k) - f(x_{k+1})). \end{aligned}$$

Zbrajanjem gornjeg izraza i Teorema 13, uz $\frac{1-\tau}{\tau} = \alpha L$ dokazujemo tvrdnju. ■

Preostaje nam još ocijeniti pogrešku.

Teorem 15. *Neka je Θ gornja ocjena vrijednosti $d(x_0, x^*)$, gdje je x_0 početna aproksimacija, a x^* točka minimuma. Ako je $\bar{x} = \frac{1}{k}\sum_{i=0}^{k-1} x_i$ aritmetička sredina prvih k iteracija pojednostavljene uparene metode, te $r \geq 0$ takav da $f(x_0) - f(x^*) \leq r$, tada imamo: $f(\bar{x}) - f(x^*) \leq \frac{2\sqrt{L\Theta r}}{k}$. Nadalje, ε-aproksimaciju točke x^* postižemo u $\mathcal{O}(\sqrt{\frac{L\Theta}{\varepsilon}})$ iteracija.*

Dokaz. Neka je $\bar{x} = \frac{1}{k}\sum_{i=0}^{k-1} x_i$ suma prvih k iteracija pojednostavljene uparene metode. U iskazu Teorema 14 uzmimo da je $u = x^*$. Korištenjem Jensenove nejednakosti imamo:

$$
\begin{aligned}
\alpha k(f(\bar{x})-f(x^*)) &= \alpha k\left(f\left(\sum_{i=0}^{k-1}\frac{x_i}{k}\right)-f(x^*)\right)\leq \alpha k\left(\sum_{i=0}^{k-1}\frac{1}{k}f(x_i)-f(x^*)\right)\\
&= \alpha\sum_{i=0}^{k-1}(f(x_i)-f(x^*))\leq \alpha\sum_{i=0}^{k-1}\langle\nabla f(x_i),x_i-x^*\rangle\\
&\leq \alpha^2L(f(y_0)-f(y_k))+d(x_0,x^*)-d(z_k,x^*)\leq\alpha^2Ld+\Theta,
\end{aligned}
$$

gdje je d neka gornja međa za $f(y_0)-f(y_k)$. Iz izbora $\alpha=\sqrt{\frac{\Theta}{Ld}}$ slijedi $f(\bar{x})-f(x^*)\leq\frac{2\sqrt{\Theta Ld}}{k}$.

Drugim riječima, u $k\geq 4\sqrt{\frac{L\Theta}{d}}$ koraka možemo doći do aproksimacije $\bar{x}$ takve da je

$f(\bar{x})-f(x^*)\leq\frac{d}{2}$. Ako ovaj postupak ponovno pokrenemo nekoliko puta, prepolavljajući udaljenost do optimuma, dobivamo ε-aproksimaciju rješenja u

$k=\mathcal{O}\left(\sqrt{\frac{L\Theta}{\varepsilon}}+\sqrt{\frac{L\Theta}{2\varepsilon}}+\sqrt{\frac{L\Theta}{4\varepsilon}}+\dots\right)=\mathcal{O}\left(\sqrt{\frac{L\Theta}{\varepsilon}}\right)$ koraka. Time je tvrdnja dokazana. ■

Iz posljednjeg dokaza možemo primijetiti kako se duljina koraka α povećava svakim ponovnim pokretanjem algoritma, dok τ pada. To znači da što smo bliže optimumu gradijentnoj metodi dajemo veću težinu, što možda nije intuitivno. Osim toga, kako bismo dodijelili vrijednosti α i τ, moramo dobro ocijeniti vrijednost Θ, te dobro odabrati početnu aproksimaciju i vrijednost d. Još jedna negativna strana ovog algoritma je to što ga moramo pokretati više puta kako bismo postigli željenu preciznost. Sljedeća metoda koju ćemo konstruirati pokušat će savladati sve navedene prepreke.

Uparena metoda

Prilikom konstrukcije ove metode zadržat ćemo pretpostavke pojednostavljene uparene metode. Dodatno ćemo zahtijevati da je $S\neq\mathbb{R}^n$ konveksan i zatvoren, tj. sada ćemo ograničenja uzeti u obzir.

Iteracija uparene metode slična je iteraciji pojednostavljene iterativne metode. Ona glasi:

- postavimo 3 početne iteracije $x_0=y_0=z_0$,
- u svakoj iteraciji k računamo vrijednosti x_{k+1}, y_{k+1}, i z_{k+1},

- također, računamo α_k i τ_k,
- $x_{k+1} = \tau_k z_k + (1-\tau_k)y_k$,
- $y_{k+1} = \mathrm{Grad}(x_{k+1})$,
- $z_{k+1} = \mathrm{Mirr}(\alpha_k, x_{k+1}, \nabla f)$.

Prvo ćemo izraziti τ_k pomoću α_k. Nadalje, α_k ćemo birati tako da τ_k bude iz intervala $(0,1]$, što će nam osigurati da iteracije ostanu unutar konveksnog skupa S. Sjetimo se, u pojednostavljenoj uparenoj metodi duljina koraka τ_k bila je jednaka $\frac{1}{\alpha_{k+1}L+1}$. U uparenoj metodi, duljinu koraka τ_k biramo na sličan način, o čemu govori sljedeća tvrdnja.

Teorem 16. *Ukoliko je $\tau_k = \frac{1}{\alpha_{k+1}L}$, tada za svaki u iz S vrijedi:*

$$\begin{aligned}\alpha_{k+1}\langle \nabla f(x_{k+1}), z_k - u\rangle &\leq \alpha_{k+1}^2 L\mathrm{Prog}(x_{k+1}) + d(z_k,u) - d(z_{k+1},u)\\ &\leq \alpha_{k+1}^2 L(f(x_{k+1}) - f(y_{k+1})) + d(z_k,u) - d(z_{k+1},u).\end{aligned}$$

Dokaz. Ekvivalentno dokazu Teorema 10, uz promjenu oznaka, imamo:

$$\begin{aligned}\alpha_{k+1}\langle \nabla f(x_{k+1}), z_k - u\rangle &= \langle \alpha_{k+1}\nabla f(x_{k+1}), z_k - z_{k+1}\rangle + \langle \alpha_{k+1}\nabla f(x_{k+1}), z_{k+1} - u\rangle\\ &\leq \langle \alpha_{k+1}\nabla f(x_{k+1}), z_k - z_{k+1}\rangle + \langle -\partial_2 d(z_k, z_{k+1}), z_{k+1} - u\rangle\\ &= \langle \alpha_{k+1}\nabla f(x_{k+1}), z_k - z_{k+1}\rangle + d(z_k,u) - d(z_{k+1},u) - d(z_k, z_{k+1})\\ &\leq (\langle \alpha_{k+1}\nabla f(x_{k+1}), z_k - z_{k+1}\rangle - \frac{1}{2}\|z_k - z_{k+1}\|^2) +\\ &\quad + (d(z_k,u) - d(z_{k+1},u)).\end{aligned}$$

Neka je sad $v = \tau_k z_{k+1} + (1-\tau_k)y_k \in S$. Tada vrijedi:

$$x_{k+1} - v = (\tau_k z_k + (1-\tau_k y_k)) - v = \tau_k(z_k - z_{k+1}).$$

Koristeći prethodnu jednakost i definiciju od $\mathrm{Prog}(x_{k+1})$ imamo:

$$
\begin{aligned}
& \langle \alpha_{k+1} \nabla f(x_{k+1}), z_k - z_{k+1} \rangle - \frac{1}{2} \|z_k - z_{k+1}\|^2 \\
= \ & \frac{\alpha_{k+1}}{\tau_k} \langle \nabla f(x_{k+1}), x_{k+1} - v \rangle - \frac{1}{2\tau_k^2} \|x_{k+1} - v\|^2 \\
= \ & \alpha_{k+1}^2 L (\langle \nabla f(x_{k+1}), x_{k+1} - v \rangle - \frac{L}{2} \|x_{k+1} - v\|^2) \\
\leq \ & \alpha_{k+1}^2 L \cdot \mathrm{Prog}(x_{k+1}),
\end{aligned}
$$

iz čega slijedi prva nejednakost tvrdnje teorema. Druga nejednakost slijedi iz analize metode gradijentnog spusta za koju vrijedi $f(x_{k+1}) - f(y_{k+1}) \geq \mathrm{Prog}(x_{k+1})$. ∎

Sljedeća nam tvrdnja donosi ocjenu pogreške iteracije uparene metode. Također, moći ćemo odrediti duljine koraka α_k.

Teorem 17. *Neka je $u \in S$. Vrijedi:*

$$\alpha_{k+1}^2 L f(y_{k+1}) - (\alpha_{k+1}^2 L - \alpha_{k+1}) f(y_k) + d(z_{k+1}, u) - d(z_k, u) \leq \alpha_{k+1} f(u).$$

Dokaz. Koristeći prethodni teorem, konveksnost od f, definicije od x_k, y_k i z_k, te izbor τ_k imamo:

$$
\begin{aligned}
& \alpha_{k+1}(f(x_{k+1}) - f(u)) \\
\leq \ & \alpha_{k+1} \langle \nabla f(x_{k+1}), x_{k+1} - u \rangle \\
= \ & \alpha_{k+1} \langle \nabla f(x_{k+1}), x_{k+1} - z_k \rangle + \alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - u \rangle \\
= \ & \frac{(1-\tau_k)\alpha_{k+1}}{\tau_k} \langle \nabla f(x_{k+1}), y_k - x_{k+1} \rangle + \alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - u \rangle \\
\leq \ & \frac{(1-\tau_k)\alpha_{k+1}}{\tau_k} (f(y_k) - f(x_{k+1})) + \alpha_{k+1} \langle \nabla f(x_{k+1}), z_k - u \rangle \\
\leq \ & \frac{(1-\tau_k)\alpha_{k+1}}{\tau_k} (f(y_k) - f(x_{k+1})) + \alpha_{k+1}^2 L (f(x_{k+1}) - f(y_{k+1})) + d(z_k, u) - d(z_{k+1}, u) \\
= \ & (\alpha_{k+1}^2 L - \alpha_{k+1}) f(y_k) - \alpha_{k+1}^2 L f(y_{k+1}) + \alpha_{k+1} f(x_{k+1}) + d(z_k, u) - d(z_{k+1}, u).
\end{aligned}
$$

Time je tvrdnja dokazana. ∎

Kako bismo mogli dokazati sljedeći teorem koji nam osigurava konvergenciju uparene metode, trebamo odabrati α_k tako da je $\alpha_k^2 L \approx \alpha_{k+1}^2 L - \alpha_{k+1}$. Izaberimo $\alpha_k = \frac{k+1}{2L}$. Tada je $\alpha_k^2 L = \alpha_{k+1}^2 L - \alpha_{k+1} + \frac{1}{4L}$.

Teorem 18. *Neka je Θ gornja ocjena vrijednosti $d(x_0, x^*)$, gdje je x_0 početna aproksimacija, a x^* optimum. U k-toj iteraciji uparene metode imamo: $f(y_k) - f(x^*) \leq \frac{4\Theta L}{(k+1)^2}$.*

Dokaz. Teleskopiranjem prethodnog teorema za $i = 0, \dots, k$. Dobivamo:

$$\alpha_k^2 L f(y_k) + \sum_{i=1}^{k-1} \frac{1}{4L} f(y_k) + d(z_k, u) - d(z_0, u) \leq \sum_{i=0}^{k} \alpha_k f(u).$$

Ukoliko uzmemo da je $u = x^*$, $f(y_i) \geq f(x^*)$, $d(z_k, u) \geq 0$, te $d(z_0, u) \leq \Theta$ imamo:

$$\begin{aligned} \frac{(k+1)^2}{4L^2} L f(y_k) &\leq \left(\tfrac{k(k+3)}{4L} - \tfrac{k-1}{4L}\right) f(x^*) + \Theta, \\ f(y_k) - f(x^*) &\leq \frac{4\Theta L}{(k+1)^2}. \end{aligned}$$

Time je tvrdnja dokazana. ■

U nastavku slijedi algoritam uparene metode.

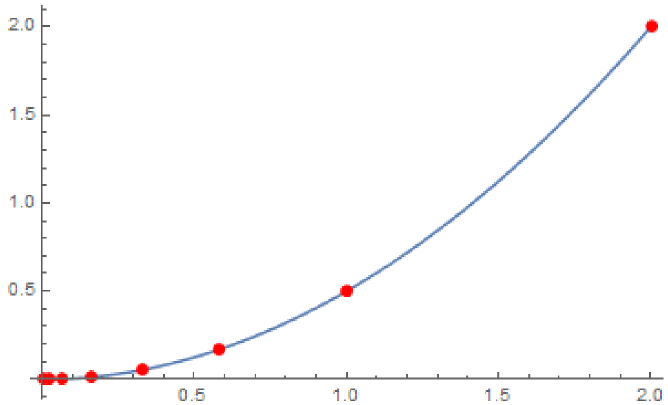
Algorithm 19.

Ulaz: funkcija f, x_0 početna aproksimacija, k broj koraka
Izlaz: y_k za koji $f(y_k) - f(x^*) \leq \frac{4\Theta L}{(k+1)^2}$.

UparenaMetoda(f, x_0, k)
1. $y_0 = x_0$, $z_0 = x_0$
2. **za** $i = 0, \ldots, k$ **radi**
3. $\tau_i = \frac{2}{i+2}$
4. $\alpha_{i+1} = \frac{i+1}{2L}$
5. $x_{i+1} = \tau_i z_i + (1 - \tau) y_i$
6. $y_{i+1} = \mathrm{Grad}(x_{i+1})$
7. $z_{i+1} = \mathrm{Mirr}(\alpha_{i+1}, x_{i+1}, \nabla f)$
8. **vrati** y_k

U sljedećem primjeru ćemo vidjeti način na koji djeluje uparena metoda.

Primjer 20. *Neka je funkcija f jednaka onoj u Primjeru 12. Uz jednaku početnu aproksimaciju i ograničenja na varijable uparenom metodom dobivamo sljedeće aproksimacije.*



Slika 3: Aproksimacije minimuma dobivene uparenom metodom

Na Slici 3 vidimo kako uparena metoda zadržava dobra svojstva gradijentne metode na prvim iteracijama, te zrcalne metode na kasnijim iteracijama.

Tablica 1 pokazuje aproksimacije minimuma dobivene gradijentnom, zrcalnom i uparenom metodom iz početne aproksimacije $x = 5$. Algoritam je stao kada je postignuta točnost $5 \cdot 10^{-4}$.

Zaključak

Ukratko smo, s teorijskog aspekta, analizirali algoritme gradijentnog i zrcalnog spusta. Utvrdili smo kako je, zbog svojstava ove dvije metode, njihovo uparivanje prirodno, te kako ono rezultira teorijski bržom metodom, takozvanom uparenom metodom. Prvo smo, na intuitivan način, konstruirali pojednostavljenu verziju uparene metode s konstantnom duljinom koraka i bez skupa ograničenja, a potom smo uveli varijabilne duljine koraka i konveksne skupove ograničenja.

Tablica 1: Iteracije gradijentne (G), zrcalne (Z) i uparene metode (U)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
G	2.5	1.25	0.625	0.3125	0.1563	0.07813	0.03906	0.01953	0.009766	0.004883	0.00244	0.00122	0.00061	0.00031
Z	4.5146	3.638	2.5784	1.5771	0.8115	0.3388	0.1085	0.02424	0.003059	0.00009				
U	2.5	1.6667	0.8333	0.3333	0.1111	0.03175	0.007937	0.001764	0.00035					

Bibliografija

[1] Z. Allen-Zhu, L. Orecchia, *Using optimization to break the epsilon barrier: a faster and simpler width-independent algorithm for solving positive linear programms in parallel*, ArXiv e-prints, abs/1407.1925, 2014.

[2] Z. Allen-Zhu, L. Orecchia, *Linear coupling: an ultimate unification of gradient and mirror descent*, ArXiv e-prints, abs/1407.1537, 2015.

[3] A. Ben-Tal, A.Nemirovski, *Lectures on modern convex optimization*, Society for industrial and applied mathematics, 2013.

[4] S. Boyd, L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[5] J. A. Kelner, Y. T. Lee, L. Orecchia, A. Sidford, *An almost linear-time algorithm for approximate max flow in undirected graphs, and its multicommodity generalizations*, Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms - SODA `14, 2014.

[6] Y. Nesterov, *Introductory lectures on convex programming volume: a basic course*, Kluwer Academic Publishers, 2004.

[7] Y. Nesterov, *Primal-dual subgradient methods for convex problems*, Mathematical programming, 120(2007.), 221-259.

ISSN 1334-6083