همگرایی سرتاسری
گرادیان مختصاتی_ کاهشی احتمالاتی
برای بهینه‌سازی غیرمحدب

Ziang Chen Yingzhou Li Jianfeng Lu

Duke University, USA
Fudan University, Shanghai

سمینار بهینه‌سازی

۲ بهمن ۱۴۰۲

فهرست

مقدمه

برای یک مسئله‌ی بهینه‌سازی، روش‌های عددی بر پایه‌ی گرادیان، در صورت همگرایی، ممکن است به نقاط زیر همگرا شویم:

- مینیمم سرتاسری
- مینیمم موضعی
- **نقاط زینی**

مقدمه

در این مقاله نشان‌داده شده است که:

با **روش گرادیان کاهش مختصاتی تصادفی**، با شروع از نقطه‌ای غیر زینی اکید، با احتمال 1 به **نقطه‌ای زینی اکید** همگرا نمی‌شویم!
بعلاوه اگر تمامی نقاط زینی مسئله، **نقاط زینی اکید** باشند، حتما به نقطه‌ی **مینیمم موضعی** همگرا خواهیم شد.

هدف

با شروع از هر نقطه غیرزینی اکید،
احتمال همگرایی به نقطه‌ای زینی اکید، صفر است!

الگوریتم در نقاط زینی اکید گیر نمی‌کند!

تعاریف

نقطه بحرانی

$$\mathrm{Crit}(f) := \{x \in \mathbb{R}^d \colon \nabla f(x) = 0\}$$

نقاط زینی

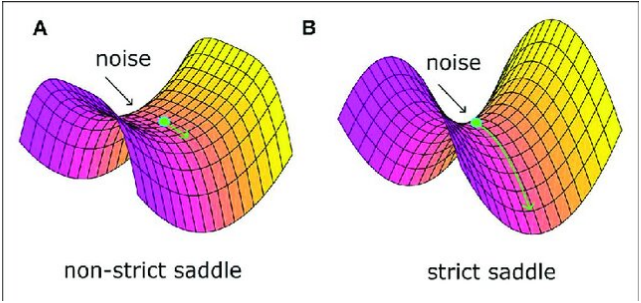
$x \in \mathrm{Crit}(f)$ یک نقطه‌ی زینی است هرگاه در هر همسایگی U آن، نقاط $x_1, x_2 \in U \setminus \{x\}$ یافت شوند که

$$f(x_1) \le f(x) \le f(x_2)$$

تعاریف

نقاط زینی اکید

$$\mathrm{Crit}_s(f) := \{x \in \mathbb{R}^d \colon \nabla f(x) = 0,\ \lambda_{\min}(H_f(x)) < 0\}$$



شکل: نقطه زینی و نقطه زینی اکید

طرح مسئله

مسئله‌ی بهینه‌سازی زیر را برای تابع هموار $f : \mathbb{R}^d \to \mathbb{R}$ در نظر بگیریم.

$$\min_{x \in \mathbb{R}^d} f(x)$$

الگوریتم گرایان کاهش مختصاتی تصادفی

الگوریتم ۱ گرادیان کاهشی با مختصات تصادفی

ورودی: $x_0 \in \mathbb{R}^d$ و $t = 0$

خروجی: مختصات مکانی و زمانی مینیمم موضعی تابع f

تا وقتی همگرایی بر قرار نشده:

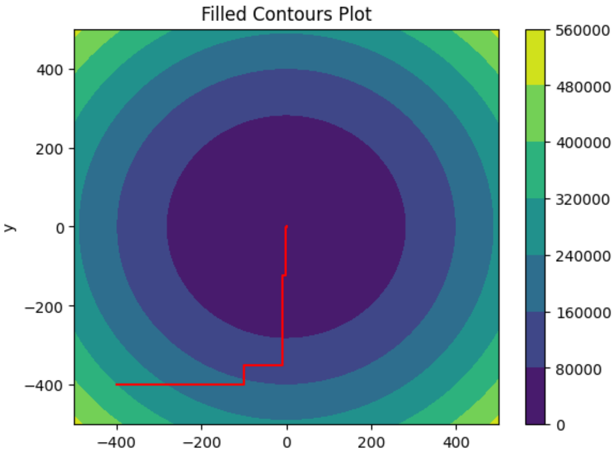
به طور تصادفی i_t را از مجموعه $\{1, 2, \ldots d\}$ ، با تابع توزیع یکنواخت، انتخاب کنید.

طول گام α_t را به طور تصادفی از بازه‌ی $[\alpha_{min}, \alpha_{max}]$ ، با تابع توزیع یکنواخت، انتخاب کنید.

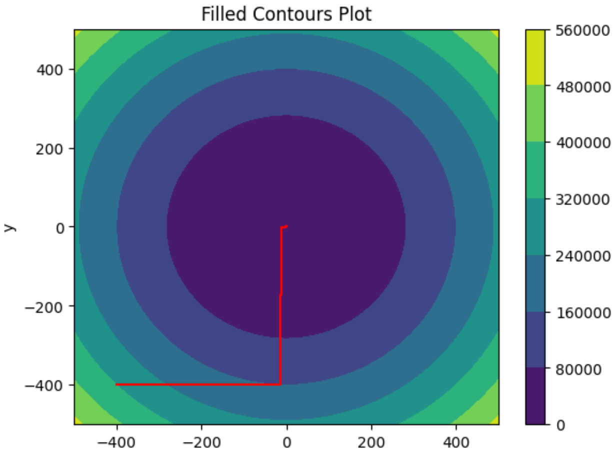
$x_t - \alpha_t e_{i_t} \partial_{i_t} f(x_t) \to x_{t+1}$

$t + 1 \to t$

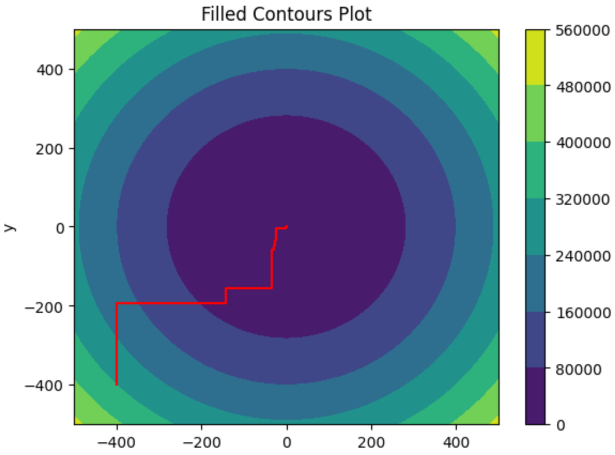
الگوریتم گرایان کاهش مختصاتی تصادفی



الگوریتم گرایان کاهش مختصاتی تصادفی



الگوریتم گرایان کاهش مختصاتی تصادفی



هدف

با شروع از هر نقطه غیرزینی اکید،
احتمال همگرایی به نقطه‌ای زینی اکید، صفر است!

الگوریتم در نقاط زینی اکید گیر نمی‌کند!

ابزار

- **سیستم‌های دینامیکی تصادفی**

تعریف سیستم دینامیکی تصادفی متناظر با مسئله

- $(\Omega, \mathcal{F}, \mathbb{P})$: فضای احتمالی که از آن مختصات و طول گام‌ها انتخاب می‌شوند.

$$\omega = (\pi_0(\omega), \pi_1(\omega), \dots) \in \Omega$$
$$\pi_t(\omega) = (i_t, \alpha_t) \in \{1, 2, \dots, d\} \times [\alpha_{\min}, \alpha_{\max}]$$

- $\theta(t)\omega = \tau^t\omega := (\pi_t(\omega), \pi_{t+1}(\omega), \dots)$: دینامیک سیستم
- $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$: فضای اندازه‌پذیری که حکم دامنه‌ی f را دارد و همگرایی الگوریتم در آن معنا دارد.

تعریف سیستم دینامیکی تصادفی متناظر با مسئله

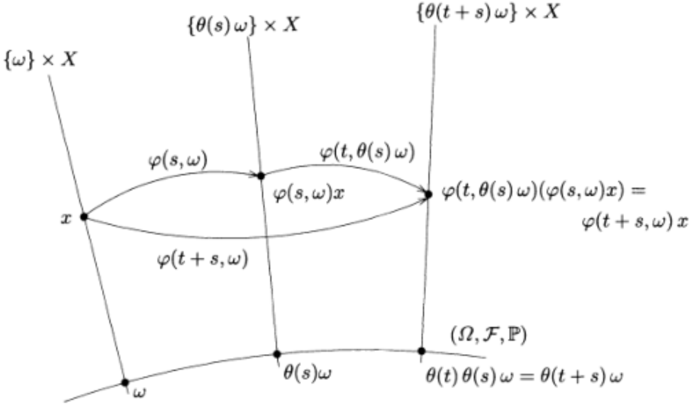
- $\varphi\colon \mathbb{N}\times\Omega\times\mathbb{R}^d \to \mathbb{R}^d$: ارتباط بین فضاهای تعریف شده - الگوریتم مسئله

$$\begin{cases} \phi(\omega)\colon \mathbb{R}^d & \to \mathbb{R}^d \\ \qquad\quad x & \to x - \alpha e_i e_i^T \nabla f(x) \end{cases}$$

برای $(i,\alpha) = \pi_0(\omega)$ در نظر بگیریم. با استفاده از این تابع، دینامیک احتمالاتی سیستم را به این صورت تعریف می‌کنیم:

$$\begin{cases} \varphi\colon \mathbb{N}\times\Omega\times\mathbb{R}^d \to \mathbb{R}^d \\ \varphi(0,\omega,x) := x \\ \varphi(t,\omega,x) := \phi(\tau^{t-1}\omega)\circ\cdots\circ\phi(\tau\omega)\circ\phi(\omega)x; \quad \forall t\geq 1 \end{cases}$$

تعریف سیستم دینامیکی تصادفی متناظر با مسئله



خطی‌سازی سیستم

سیستم خطی شده

$$\Phi(t,\omega) = A(\tau^{t-1}\omega)\ldots A(\tau\omega)A(\omega)$$
$$A(\omega) = I - \alpha e_i e_i^T H(f); \quad (i,\alpha) = \pi_0(\omega)$$

تعریف

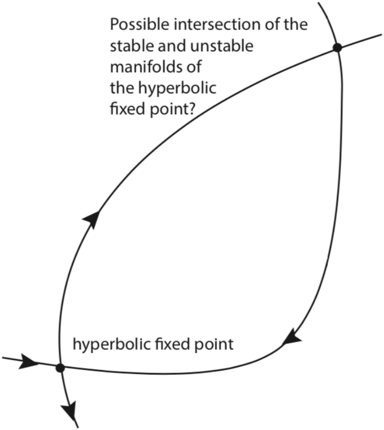
خمینه‌ی پایدار

$$W^s(x^*) := \{x_0 \in \mathbb{R}^d : \lim_{t\to\infty} x_t = x^* \quad \text{به صورت نمایی}\}$$

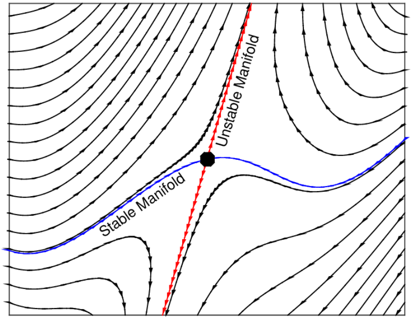
خمینه‌ی ناپایدار

$$W^u(x^*) := \{x_0 \in \mathbb{R}^d : \exists \epsilon > 0 \text{ s.t. } ||x_t|| \geq e^{\epsilon t} \text{ when } t \to \infty\}$$

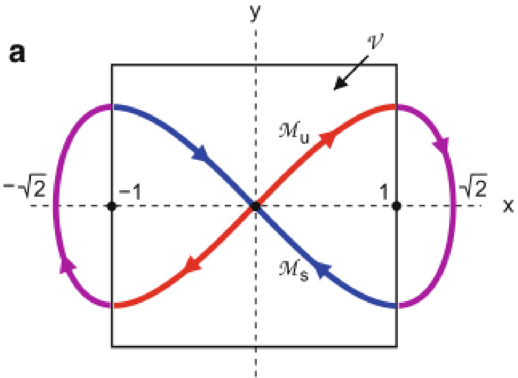
خمینه‌ی پایدار و ناپایدار



خمینه‌ی پایدار و ناپایدار



آیا تحلیل موضعی کافی است؟



خطی‌سازی سیستم

سیستم خطی شده

$$\Phi(t,\omega) = A(\tau^{t-1}\omega)\dots A(\tau\omega)A(\omega)$$
$$A(\omega) = I - \alpha e_i e_i^T H(f); \quad (i,\alpha) = \pi_0(\omega)$$

- اگر $\{x^*\} \neq W^u(x^*)$ آنگاه $d > \dim(W^s(x^*))$ پس

$$\mathbb{P}(W^s(x^*)) = 0$$

فرضیات

فرض اول

- $f \in C^2(\mathbb{R}^d)$
- $H(f)$ به طور یکنواخت کراندار است یعنی:

$$\exists M > 0 \text{ s.t. } \forall x \in \mathbb{R}^d : ||H(f)x|| \leq M$$

فرض دوم

- برای تمامی $x^* \in \text{Crit}_s(f)$، تمامی مقادیر ویژه ماتریس هستیان در x^* ناصفر باشند.
  $\leftarrow$ نقاط زینی اکید منزوی هستند

فرض سوم

- $\alpha_{\max} < \frac{1}{M}$ $\leftarrow$ وجود خمینه‌ی ناپایدار

نتایج

نتیجه اول

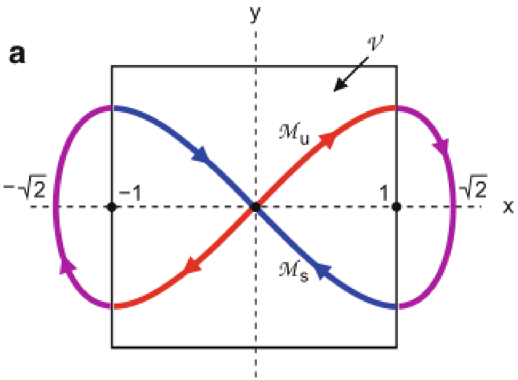
با فرضیات مناسب بیان شده، بزرگ‌ترین نمای لیاپانوف سیستم خطی، مثبت است.

خمینه‌ی ناپایدار نابدیهی است!

- در نتیجه‌ی آن،

$$\dim(W^u) > 0$$

آیا تحلیل موضعی سیستم کافی نیست!



تعمیم به حالت غیرخطی

قضیه

با تمامی مفروضات بالا $\epsilon^* \in (0, \frac{1}{6})$ وجود دارد که برای هر $\epsilon \in (0, \epsilon^*)$ گام زمانی‌ای مانند T^* وجود دارد که برای $T \geq T^*$ و هر $t \in \mathbb{N}$، عددی مانند η وجود دارد که با احتمال $1 - 4\epsilon$

$$||x_{t+T}|| \geq e^{\eta}||x_t||$$

نتایج

$$\Omega(x^*, x_0) := \{\omega \in \Omega \colon \lim_{t\to\infty} x_t = x^*\}$$

$$\Omega(\mathrm{Crit}_s(f), x_0) := \bigcup_{x^* \in \mathrm{Crit}_s(f)} \Omega(x^*, x_0)$$

نتیجه اصلی

با فرضیات مناسب بیان شده، برای هر $x_0 \in \mathbb{R}^d \setminus \mathrm{Crit}_s(f)$ داریم

$$\mathbb{P}(\Omega(\mathrm{Crit}_s(f), x_0)) = 0$$

طرح اثبات

- برای هر $x^* \in \mathrm{Crit}_s(f)$: $\quad \mathbb{P}(\Omega(x^*, x_0)) = 0$
- $\mathrm{Crit}_s(f)$ شمارا است. چرا که هر نقطه آن منزوی است.
- $\mathbb{P}(\Omega(\mathrm{Crit}_s(f), x_0)) \leq \sum_{x^* \in \mathrm{Crit}_s(f)} \mathbb{P}(\Omega(x^*, x_0)) = 0$

نتایج

نتیجه سوم

با تمامی مفروضات بالا، و فرض اضافی منزوی بودن تمامی نقاط بحرانی با $\mathrm{Crit}_s(f) \setminus \mathrm{Crit}(f) \ni x^*$ به نقطه‌ی $\mathrm{Crit}_s(f)$ با احتمال 1 $\{x_t\}$ دنباله‌ی $x_0 \in \mathbb{R}^d \setminus \mathrm{Crit}_s(f)$،

با فرض اکید بودن تمامی نقاط زینی سیستم، با احتمال 1 به یک نقطه مینیمم موضعی همگرا خواهیم بود.

طرح اثبات

- با احتمال 1، مجموعه نقاط حدی دنباله‌ی $\{x_t\}$ زیرمجموعه‌ای از نقاط بحرانی است.
- با فرض منزوی بودن نقاط بحرانی، با احتمال 1، دنباله‌ی $\{x_t\}$ همگرا بوده و حد آن یک نقطه‌ی بحرانی است.
- با نتیجه‌ی اصلی، می‌دانیم این نقطه با احتمال 1 یک نقطه‌ی زینی اکید نخواهد بود. پس نتیجه برقرار خواهد بود.

منابع

L. Arnold, C. K. Jones, K. Mischaikow, G. Raugel, and L. Arnold.
Random dynamical systems.
Springer, 1995.

J. Sherman and W. J. Morrison, “Adjustment of an inverse matrix corresponding to a change in one element of a given matrix,” *The Annals of Mathematical Statistics*, vol.21, no.1, pp.124–127, 1950.

M. S. Bartlett, “An inverse matrix adjustment arising in discriminant analysis,” *The Ann of Mathematical Statistics*, vol.22, no.1, pp.107–111, 1951.

S. J. Wright, “Coordinate descent algorithms,” *Mathematical programming*, vol.151, no.1, pp.3–34, 2015.

پرسش و پاسخ

پرسش...