これまで，最適化問題の解がみたす条件をみてきた。

解きたい問題は ①〜③．

→ $\underbrace{\nabla f(x)}$ $=0$ をみたす x を求めればOK．

↑ 直接解くのは難しい ⇒ 反復法

↑

∇f の計算できるようになった

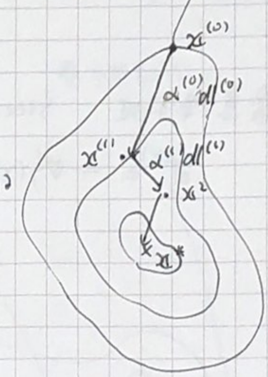
等式制約をみたす x に対して $\nabla f(x)$ を計算 ← 随伴変数法でできる

→ 不等式制約さえ考えればOK．

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{s.t.} \quad g_i(x) \le 0 \quad \text{for } i = 1 \sim m$$

~~(はじめく 無制約問題でそえる)~~

f の等高線



反復法

(0) $x^{(0)}$ を $\left(g_i(x^{(0)}) \le 0 \text{ をみたすように} \right)$ 与える，$k=0$ とする

(1) 探索方向 $d^{(k)}$ を「何らかの方法で」計算．
とステップサイズ $\alpha^{(k)} > 0$

(2) $x^{(k+1)} = x^{(k)} + \underbrace{\alpha^{(k)}}_{\text{ステップサイズ}} d^{(k)}$ と x を更新．

(3) 終了条件をみたしたら $x^{(k+1)} = x^*$ とする
そうでなければ $k \leftarrow k+1$ として (1) にもどる

(終了条件) ・$f(x^{(\cdots)})$ が十分に小さい
・$\|\nabla f(x^{(k+1)})\|$ が十分に小さい
・$\alpha^{(k+1)}$ が十分に小さい．

以降，しばらく制約がない場合を考える
(or 等式制約のみ)．

探索方向 d の決め方

→ 降下法　$f(x^{(0)}) > f(x^{(1)}) > \cdots$ となるように $x^{(k)}$ を生成.

$\underset{\downarrow 0}{\alpha^{(k)}}$ が十分に小さければ $f(x^{(k+1)})$ と $f(x^{(k)})$ の関係は

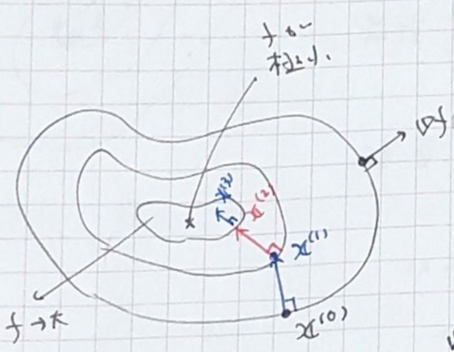
$$
\begin{aligned}
f(x^{(k+1)}) &= f(x^{(k)} + \alpha^{(k)} d^{(k)}) \\
&\approx f(x^{(k)}) + \nabla f(x^{(k)}) \cdot \alpha^{(k)} d^{(k)} \\
&= f(x^{(k)}) + \alpha^{(k)} \nabla f(x^{(k)}) \cdot d^{(k)}
\end{aligned}
$$

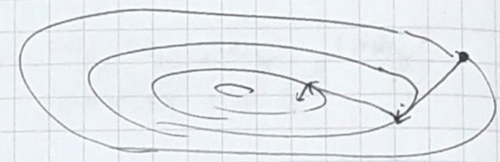
$f(x^{(k+1)}) < f(x^{(k)})$ とするためには $\boxed{\nabla f(x^{(k)}) \cdot d^{(k)} < 0}$ ……⑩ をみたす d を使えばよい

⑭ 最急降下法　Steepest decent method

$d^{(k)} = -\nabla f(x^{(k)})$ と選ぶ. → $\nabla f(x^{(k)}) \cdot d^{(k)} = -\|\nabla f(x^{(k)})\|^2 < 0$

より ⑩ は常に満たされる





おそい例

(かんたん) 一般に, 収束は "遅い" が 収束するが[illegible]

⑪ Newton 法 → $\nabla f(x) = 0$ を目指す.

$$
\frac{\partial f(x^{(k+1)})}{\partial x_i} = \frac{\partial f(x^{(k)} + \alpha^{(k)} d^{(k)})}{\partial x_i}
= \frac{\partial f(x^{(k)})}{\partial x_i} + \sum_{j=1}^{n} \underbrace{\boxed{\frac{\partial}{\partial x_j}\left(\frac{\partial f(x^{(k)})}{\partial x_i}\right)}}_{(H_f(x^{(k)}))_{ij}} \alpha^{(k)} d_j^{(k)}
$$

$\nabla f(x^{(k+1)}) = 0$

↑ 願望

$$
\begin{aligned}
\nabla f(x^{(k+1)}) &= \nabla f(x^{(k)} + \alpha^{(k)} d^{(k)}) \\
&\approx \nabla f(x^{(k)}) + \alpha^{(k)} H_f(x^{(k)}) d^{(k)} = 0.
\end{aligned}
$$

$$
\Leftrightarrow \quad \alpha^{(k)} d^{(k)} = -H_f^{-1}(x^{(k)}) \nabla f(x^{(k)}).
$$

→ 方向として $d^{(k)} = -H_f^{-1}(x^{(k)}) \nabla f(x^{(k)})$ を使ってみる.

⑩は? → $\nabla f(x^{(k)}) \cdot d^{(k)} = \nabla f(x)^T H_f^{-1}(x^{(k)}) \nabla f(x^{(k)})$ より $H_f^{-1}(x^{(k)})$ が正定値なら ⑩ をみたし

デメリット： ・降下方向が常に生成されるわけではない

・$H_f(x^{(k)})\, d^{(k)} = -\nabla f(x^{(k)})$ をとくのがたいへん，$H_f(x^{(k)})$ を計算するのもたいへん

メリット ・$x^{(0)}$ が"十分に"最適解に近ければ収束は速い．

さいきゅうこうか $d^{(k)} = -\nabla f(x^{(k)}) = -I\, \nabla f(x^{(k)})$ たんい行列

にゅーとん $d^{(k)} = -H_f^{-1}(x^{(k)})\, \nabla f(x^{(k)})$.

$I \to H_f^{-1}(x^{(k)})$ におきかえたものに見える．

I と H_f^{-1} のあいだくらいのはない？ → 準Newton法

(2) 準Newton法

余談的に [illegible].

$$B^{(k)} d^{(k)} = -\nabla f(x^{(k)})$$ をみたす $B^{(k)}$ をつかう

- $B^{(k)}$ は漸化的に定める． $B^{(0)} = I$ とか．
- $B^{(k)}$ は正定値であってほしい (2) ・対称であってほしい．
- 〃 は計算するのがかんたんであってほしい．
- 〃 は $H_f(x^{(k)})$ となんらかのイミで近似してて…ほしい．

BFGS公式 $s^{(k)} = x^{(k+1)} - x^{(k)}$

$y^{(k)} = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$ とし，

$$B^{(k+1)} = B^{(k)} - \frac{(B^{(k)} s^{(k)})(B^{(k)} s^{(k)})^T}{(s^{(k)})^T B^{(k)} s^{(k)}} + \frac{y^{(k)} (y^{(k)})^T}{(y^{(k)})^T s^{(k)}}$$

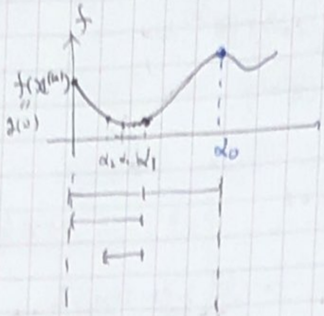
ステップサイズの決め方

(i) 厳密な直線探索

$\min_{\alpha} g(\alpha) = f(x^{(k)} + \alpha d^{(k)})$ をみたすような α を $\alpha^{(k)}$ とする

1次元の最適化問題なので解けそう

(二分法)



⓪ 適当な $\alpha_M^{(0)}$ を決める

$\alpha_m^{(0)} = 0$ とする

$k=0$ とする

① $\alpha_c^{(k)} = \dfrac{\alpha_m^{(0)} + \alpha_M^{(0)}}{2}$

② $f(\alpha_m^{(k)})$, $f(\alpha_c^{(k)})$, $f(\alpha_M^{(k)})$

$f(\alpha_m^{(k)})$ が最大　$\alpha_m^{(k+1)} = \alpha_c^{(k)}$, $\alpha_M^{(k+1)} = \alpha_M^{(k)}$ とする

$f(\alpha_c^{(k)})$ が最大 または $f(\alpha_M^{(k)})$ が最大　$\alpha_m^{(k+1)} = \alpha_m^{(k)}$, $\alpha_M^{(k+1)} = \alpha_c^{(k)}$

(①へ)

($\alpha_M^{(0)}$ の選び方が難しそう…)

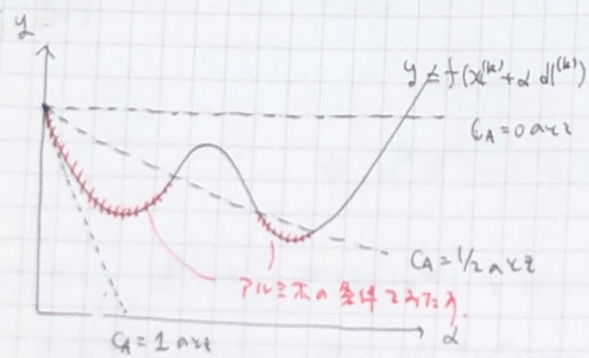
~~(分割 区間 を必要とする)~~

(ii) 非厳密な直線探索

Armijo の条件　$f(x^{(k+1)})$ が $f(x^{(k)})$ よりも「十分に」小さくなることを要求

$$f(x^{(k)} + \alpha d^{(k)}) \le f(x^{(k)}) + c_A \alpha \nabla f(x^{(k)}) \cdot d^{(k)} < f(x^{(k)})$$

をみたすように α を決める。 $0 < c_A < 1$ はパラメータ.



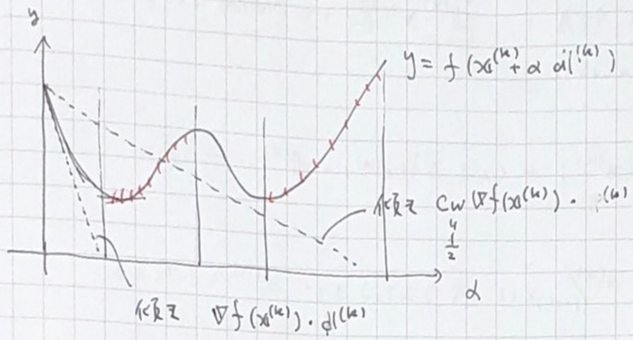
Wolfeの条件 … $\alpha^{(k)}$を十分小さくすればArmijoの条件はみたされてしまう

→あまりに小さい$\alpha^{(k)}$をえらぶといつまでたっても最適解がみつからない

↳ $\alpha^{(k)}$はあるていど大きくとりたい

$$\nabla f(x^{(k)} + \alpha d^{(k)}) \cdot d^{(k)} \geq c_W \nabla f(x^{(k)}) \cdot d^{(k)}$$

をみたすようにαをえらぶ。 $c_A \leq c_W < 1$ はパラメータ。



傾きがゆるやかになってほしい

Armijo と Wolfe の & をとる.

Why $c_A < c_W$?

$$f(x^{(k)}) \approx f(x^{(k)} + \alpha d^{(k)}) - \alpha \nabla f(x^{(k)} + \alpha d^{(k)}) \cdot d^{(k)}$$

Wolfeがみたされるならば

$$c_W \nabla f(x^{(k)}) \cdot d^{(k)} \leq \left(f(x^{(k)} + \alpha d^{(k)}) - f(x^{(k)}) \right) \frac{1}{\alpha}$$

Armijoがみたされるならば

$$\frac{1}{\alpha} \left(f(x^{(k)} + \alpha d^{(k)}) - f(x^{(k)}) \right) \leq c_A \cdot \nabla f(x^{(k)}) \cdot d^{(k)}$$

$$\Rightarrow \quad (c_W - c_A) \boxed{\nabla f(x^{(k)}) \cdot d^{(k)}} \leq 0.$$

$$\wedge\!\!|\ \ 0$$

$c_W \geq c_A$

まとめ.

$\min_{x} f(x)$ を解きたい

反復法　Input parameters $C_A \in (0,1)$, $C_W \in (C_A, 1)$　ε

⓪ $x^{(0)}$ $\alpha^{(0)}$ を適当に与える. $k=0$ とする.

① $f(x^{(k)})$, $\nabla f(x^{(k)})$ を計算し, $d^{(k)}$ を 最急 とか Newton とかで求める　探索方向

ずいはん へんろう法

アルミホ チェック

② $f(x^{(k)} + \alpha^{(k)} d^{(k)}) \leq f(x^{(k)}) + C_A \alpha \nabla f(x^{(k)}) \cdot d^{(k)}$ — Ⓐ

がみたされていれば ③へ.

そうでなければ $\alpha^{(k)} \leftarrow \alpha^{(k)}/(1+\varepsilon)$ として Ⓐをみたすまで $\alpha^{(k)}$ を小さくする

ウルフチェック

③ $\nabla f(x^{(k)} + \alpha^{(k)} d^{(k)})$ を計算

$\nabla f(x^{(k)} + \alpha^{(k)} d^{(k)}) \cdot d^{(k)} \geq C_W \nabla f(x^{(k)}) \cdot d^{(k)}$ — Ⓑ

がみたされていれば $x^{(k+1)} \leftarrow x^{(k)} + \alpha^{(k)} d^{(k)}$ とする

そうでなければ $\alpha^{(k)} \leftarrow \alpha^{(k)}/(1-\varepsilon)$ として ②にもどる

④ 終了条件をみたすなら 解を出力

そうでなければ $k \leftarrow k+1$ として ①にもどる