最適化理論

勾配法・ニュートン法

数値最適化アルゴリズム

大規模な数理計画問題を解くためには、効率の良いアルゴリズムが必要。

無制約凸問題

- 勾配法
- ニュートン法

（制約付き）凸計画問題

- 内点法

目標

- 無制約凸計画問題

$$\text{minimize } f(x)$$

を考える(f は凸関数であり、$f : \mathbb{R}^n \to \mathbb{R}$)。

- 最適化アルゴリズムは、系列
$x^{(0)}, x^{(1)}, \ldots \in \mathrm{dom} f$ を生成する。ここで、

$$f(x^{(k)}) \to p^*, \quad k \to \infty$$

である。$x^{(0)}$ を初期点と呼ぶ。

仮定

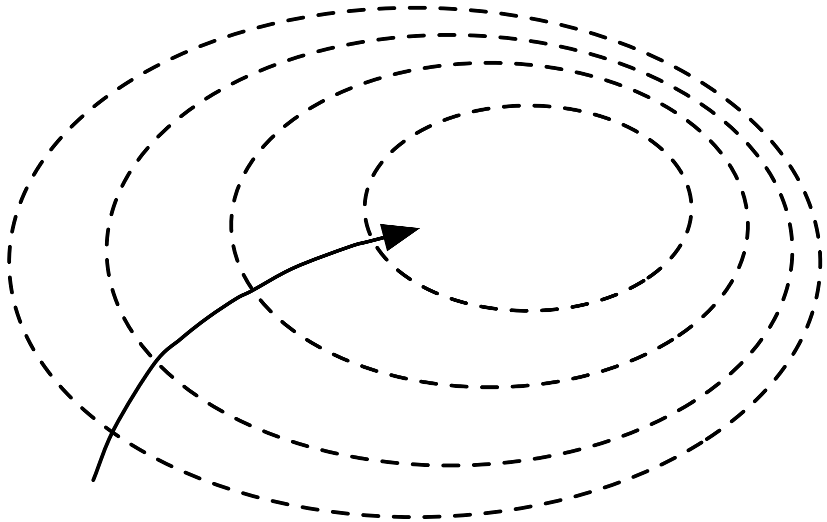
- $x^{(0)} \in \mathbf{dom}\, f$
- サブレベル集合

$$S = \{x : f(x) \leq f(x^{(0)})\}$$

が閉集合となっている。

降下法

降下法は、勾配法、ニュートン法を含む一般的な最適化原理。



- $\nabla f(x) = 0$を満たす点を探す。

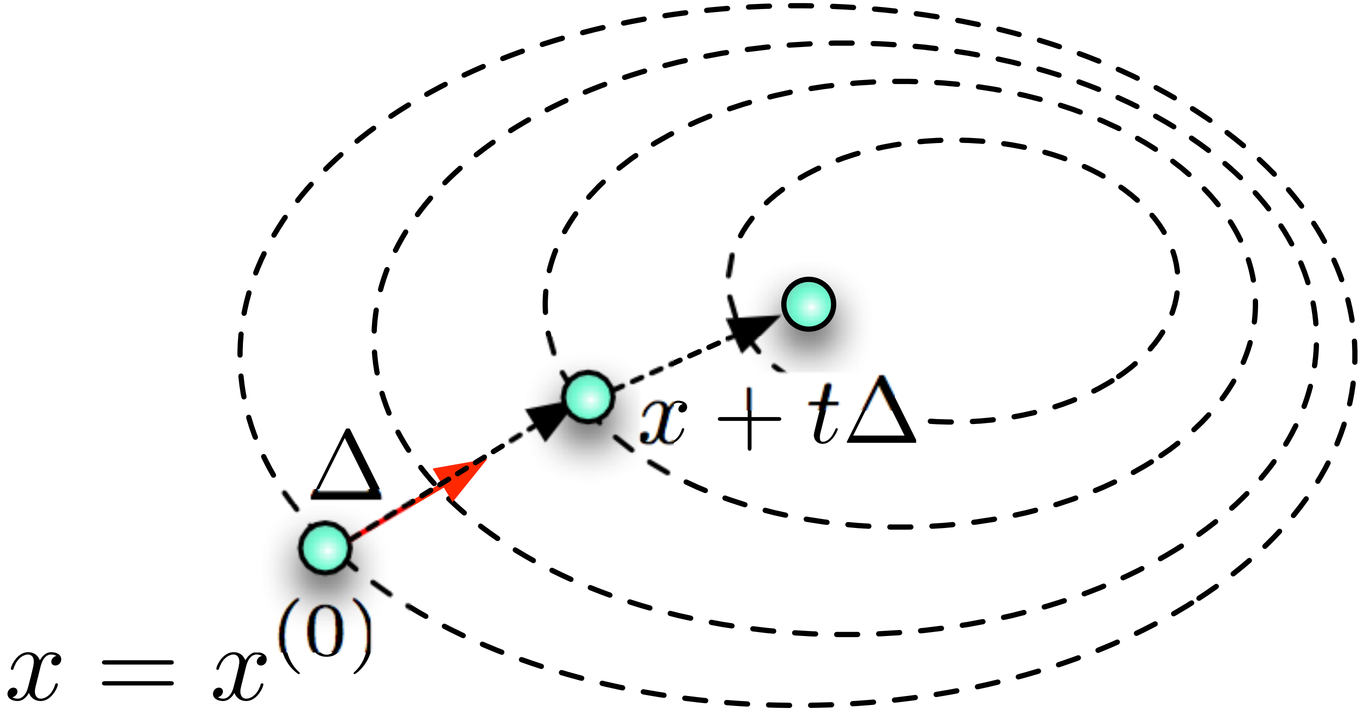
降下法(descent method)の手順

探索点を$x = x^{(0)}$としたのち、下記を繰り返す。

- xにおける探索方向ベクトルΔを求める
- 直線探索を行い、ステップ乗数$t (t > 0)$を求める。
- $x := x + t\Delta$として、探索点を更新する
- 停止条件が満たされた場合に実行終了。

注: 機械学習分野では、直線探索を行わず、定数αに対して$x := x + \alpha\Delta$と更新を行う。

降下法の実行過程



探索方向Δ

第k反復における探索方向を$\Delta^{(k)}$、ステップ乗数を$t^{(k)}$とすると

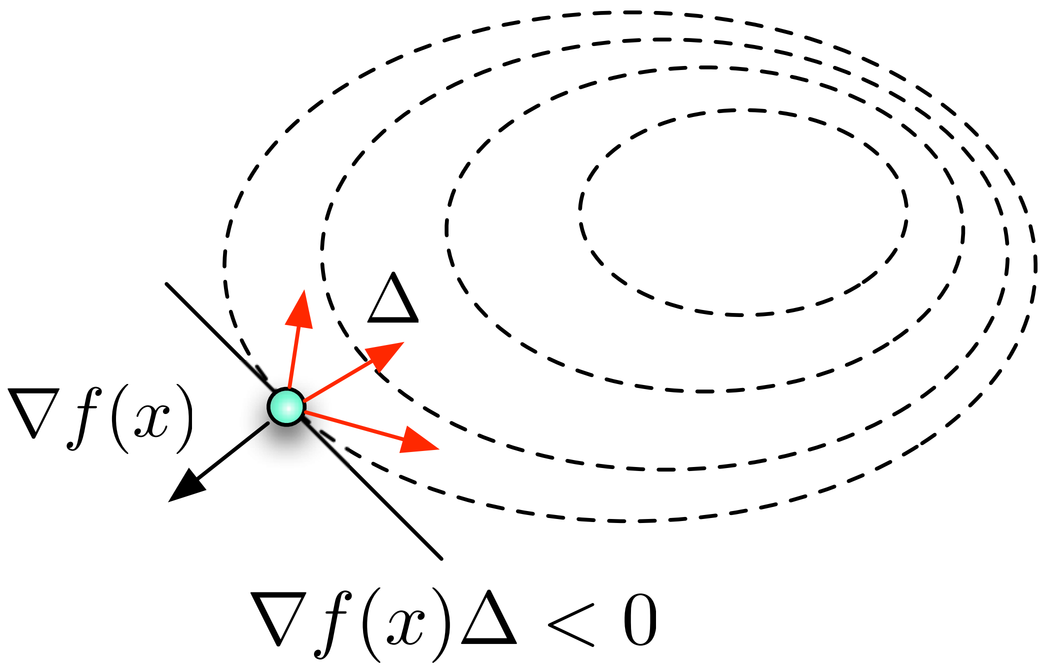
$$x^{(k+1)} = x^{(k)} + t^{(k)}\Delta^{(k)}$$

となる。

$$f(x^{(0)}) > f(x^{(1)}) > f(x^{(2)}) > \cdots$$

となるように$\Delta^{(k)}$、$t^{(k)}$を選びたい。

目的関数 f が凸関数であることより、もし、
$f(x+t\Delta) < f(x)$ が成り立つならば、
$\nabla f(x)\Delta < 0$ が成り立つ。ただし、$t > 0$ である。



直線探索 (1)

正確な直線探索

$$t = \arg\min_{t>0} f(x + t\Delta)$$

- 1次元の目的関数(tの関数)の最小化問題
- 場合によっては解析的に求まる。

直線探索 (2)

α, β を $0 < \alpha < 0.5, 0 < \beta < 1$ の範囲から選ぶ。

バックトラック探索

- $t := 1$ と設定する。
- もし、

$$f(x + t\Delta) > f(x) + \alpha t \nabla f(x)^T \Delta$$

ならば、$t := \beta t$ として、もう一度、このステップを実行。

勾配法 (gradient descent method)

勾配法は、探索方向Δとして、負勾配ベクトル$-\nabla f(x)$を利用する降下法である。

- 最適解への収束性を持つ
- 探索点は等高線に常に直交するように移動する。

勾配法の手順

探索点を$x = x^{(0)}$としたのち、下記を繰り返す。

- $\Delta = -\nabla f(x)$とする。
- $||\nabla f(x)||_2 < \epsilon$ で実行終了。
- 直線探索を行い、ステップ乗数$t(t > 0)$を求める。
- $x := x + t\Delta$として、探索点を更新する

最適解への収束

勾配法について、$x \in S$ について、t^* が存在して

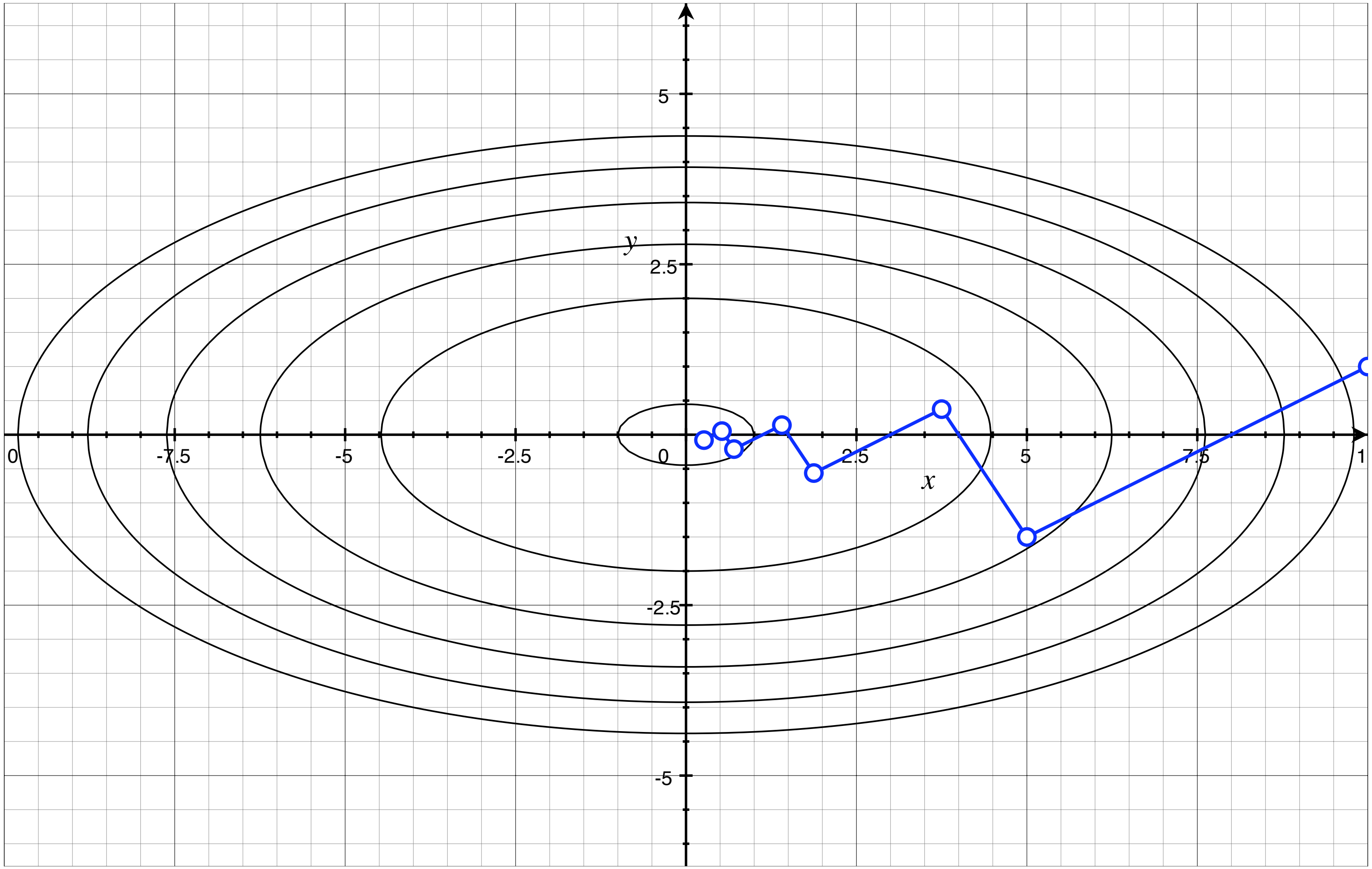
$$f(x - t^* \nabla f(x)) < f(x) - \frac{M}{2} \|\nabla f(x)\|_2^2$$

が成り立つ。$M > 0$ は f と $x^{(0)}$ から決まる定数。
したがって、

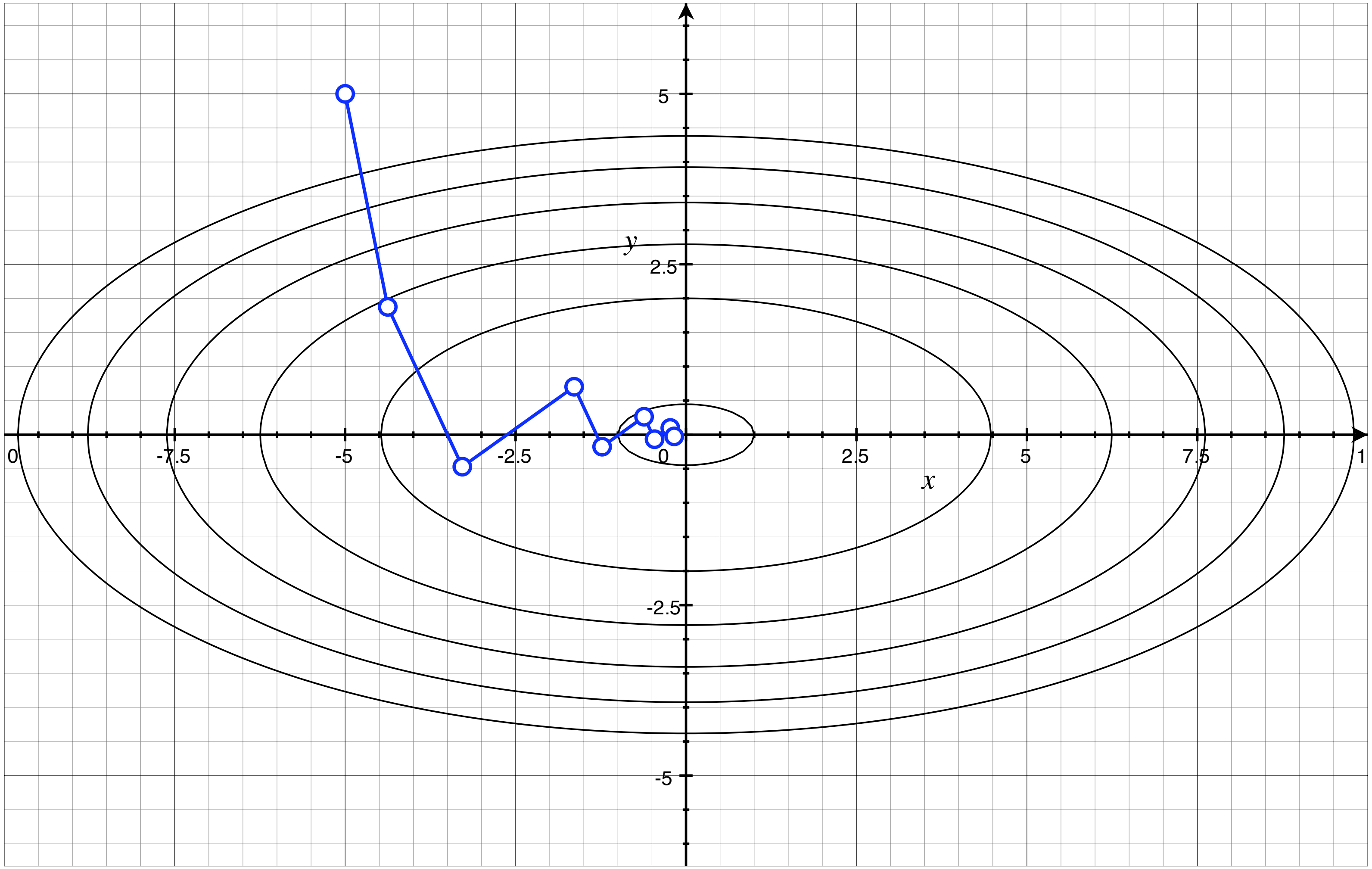
$$f(x^{(0)}) > f(x^{(1)}) > f(x^{(2)}) > \dots$$

が成り立つ。この減少数列は下に有界
$(f(x^{(k)}) \geq p^*)$ なので、p^* に収束する。

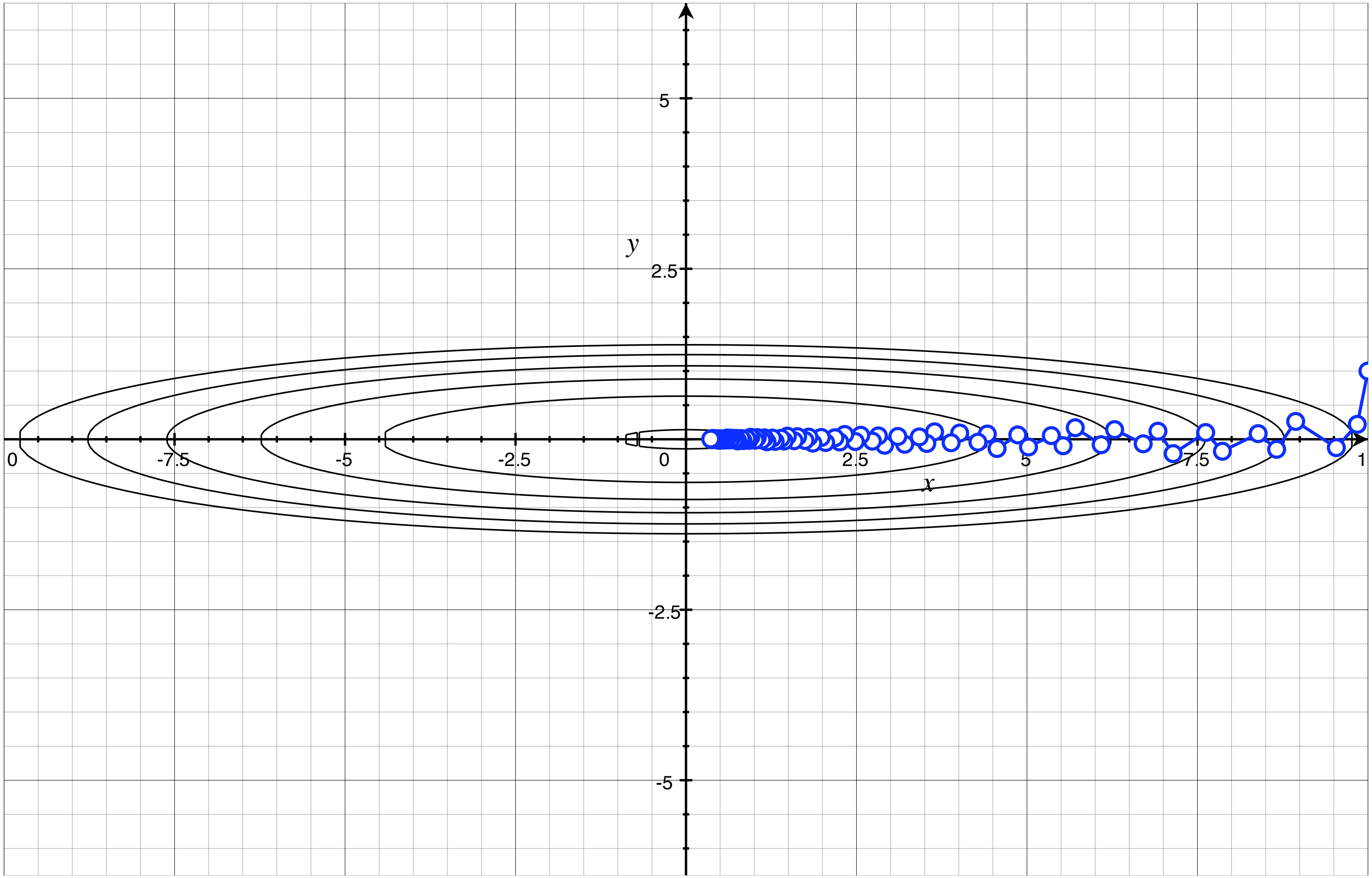
$$f(x) = x_1^2 + \theta x_2^2,\ \ \theta = 5,\ \ x^{(0)} = (10, 1)$$



$$f(x) = x_1^2 + \theta x_2^2,\ \ \theta = 5,\ \ x^{(0)} = (-5, 5)$$



$$f(x) = x_1^2 + \theta x_2^2, \;\; \theta = 50, \;\; x^{(0)} = (10, 1)$$



勾配法の収束速度

- 目的関数の形状によっては、勾配法の収束は非常に遅くなる。
- 勾配法は単純だが、多くの応用では収束が遅くなるケースが多く、実用に耐えない場合もある。
- 収束速度は目的関数のヘッセ行列の “条件数” と深く関係する。

勾配法のまとめ

- 探索方向として負勾配ベクトルを利用する降下法の一種
- 目的関数の形状によっては収束が非常に遅くなる場合がある。
- 一次収束性を持つ
- 大局的収束性を持つ

ニュートン法

ニュートン法は、目的関数の２次導関数の情報を利用することにより、より高速な最適解への収束を実現する最適化アルゴリズムである。

ニュートンステップ

ニュートン法も降下法の一種であり、探索方向として、次の"ニュートンステップ"(ニュートン方向ベクトル)を利用する。

$$\Delta_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$$

- $\nabla f(x)$: 勾配ベクトル (gradient)
- $\nabla^2 f(x)$: ヘッセ行列 (hessian matrix)

ニュートン方程式

ニュートンステップを求めるためには、次の連立一次方程式

$$Gx = -g$$

を解く必要がある(Gはヘッセ行列 ($n \times n$実行列)、gは勾配ベクトル)。

ニュートンステップの方向

凸関数 f について、ニュートンステップは降下方向を向いている：

$$\nabla f(x)^T d = -\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x) < 0$$

- $\forall x, \nabla^2 f(x) \succeq 0 \Rightarrow \forall x, \nabla^2 f(x)^{-1} \succeq 0$
- 一般の非線形関数の場合には、ニュートンステップは降下方向になっているとは限らない。

多変数関数の２次近似

多変数関数のテーラー展開を２次で打ち切ることにより、次の近似

$$f(x+v) \simeq f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$$

が得られる。

ニュートンステップの意味

$f : \mathbb{R}^n \to \mathbb{R}$を凸関数とする。$x + \Delta_{nt}$は2次局所近似

$$\hat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v$$

の極値(＝最小値)を与える。

ニュートン法の手順

ニュートン法の手順は次のとおり。

探索点を $x = x^{(0)}$ としたのち、下記を繰り返す。

- x におけるニュートンステップ Δ_{nt} を求める
- 停止条件(後述)が満たされた場合に実行終了。
- 直線探索を行い、ステップ乗数 $t(0 < t \leq 1)$ を求める。
- $x := x + t\Delta_{nt}$ として、探索点を更新する

ニュートン法の特徴

- 最適解の近くで２次収束性を持つ
- 初期点が最適解から離れている場合には、２次収束領域に到達するまでに多数の反復が必要になる可能性がある。