Introduction

Problem

$$\boxed{\min_{x \in \Omega} f(x)}$$

- f: objective function
- $\Omega \subset \mathbb{R}^n$: feasible set / constraint set

max $f(x)$ can be changed to min $-f(x)$.

Optimization

1. Modeling. Practical problems $\rightarrow$ optimization problems
2. Algorithms. Methods to solve
3. Software. Implement

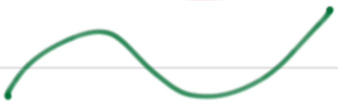
We focus on 2 and 3.

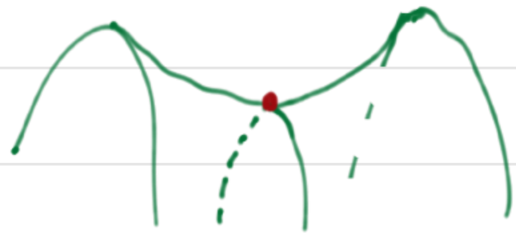
Types of Optimization Problems

easier cases

- Constrained $x \in \Omega$  v.s. Non-constrained $x \in \mathbb{R}^n$ x is free
- Convex  v.s. Non-convex 
  both f and Ω are convex
- Smooth ∇f  v.s. Non-smooth ∂f 

- 1-d $\mathbb{R}^1$ v.s. n-d $\mathbb{R}^n$





used for $\begin{cases} \text{illustration} \\ \text{motivation} \\ \text{understanding} \end{cases}$

harder due to infinite directions

reduce to one direction

Examples

1. $\min\limits_{x \in [0,5]} f(x)$, $f(x) = x^3 - 6x^2 + 9x + 1$

This is a 1-d smooth, non-convex, and constrained opt

Sol. (only outline not detailed solution).

Critical pts: $f'(x) = 0$. $x = 1$ or $x = 3$

min or max: $f''(1) < 0$ local max. $f''(3) > 0$ local max.

Question: how about the case $f''(x) = 0$?

Can we conclude the answer is $x = 3$?

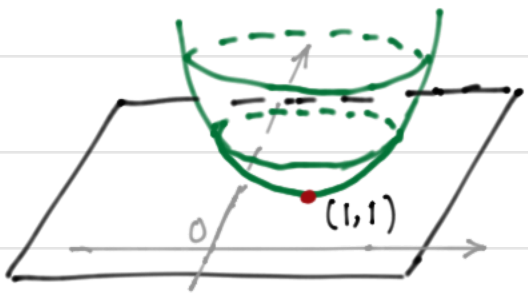
No! This is a constrained opt problem. Check $x = 3 \in \Omega = [0, 5]$ ✓.

And compare with values on boundary: $f(0)$, $f(5)$.

The final answer is $\min\limits_{x \in [0,5]} f(x) = 1$ and is acheived at $x = 0, 3$.

Question: Change the constraint $\min f(x)$ s.t. $x \in [-1, 1]$.

2 $f(x,y) = (x-1)^2 + (y-1)^2$ $\qquad \min_{(x,y)\in\mathbb{R}^2} f(x,y)$



Critical pts :

$$\begin{cases} \dfrac{\partial f}{\partial x}(x,y) = 0, & 2(x-1)=0 \Rightarrow x=1 \\ \dfrac{\partial f}{\partial y}(x,y) = 0, & 2(y-1)=0 \Rightarrow y=1. \end{cases}$$

Second order conditions.

- 1-d : $f''(x) > 0$
- 2-d : Hessian matrix $\nabla^2 f = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial y \partial x} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix}$, formally $\nabla^2 f > 0$.

For a symmetric matrix

$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$. $A > 0$ if $\forall (u,v) \in \mathbb{R}^2$, $(u,v) A \begin{pmatrix} u \\ v \end{pmatrix} > 0$.

$a_{11}u^2 + 2a_{12}uv + a_{22}v^2 > 0, \ \forall u,v$ quadratic form

Set $v=0$, $u=1$ $\rightarrow a_{11} > 0$. $\qquad \Delta = 4a_{12}^2 - 4a_{11}a_{22} < 0$

Set $u=0$, $v=1$ $\rightarrow a_{22} > 0$. $\qquad \det(A) = a_{11}a_{22} - a_{12}^2 > 0$.

2×2 symmetric $A > 0$ if a_{11} or $a_{22} > 0$ and $\det(A) > 0$.

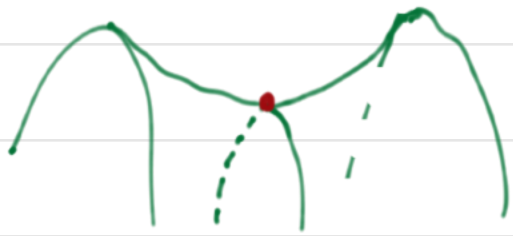
Important. For a number, either $a > 0$ or $a \leq 0$.

For a symmetric matrix, $A > 0$ or $A \not> 0$ including $\begin{cases} A \leq 0 \\ \text{Saddle point} \end{cases}$

Better to view as $A = Q^T \Lambda Q$, where $\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$ and Q consists of corresponding eigen-vectors, and $Q^T Q = I$.

$A > 0 \iff \lambda_1 > 0,\ \lambda_2 > 0$

$A \not> 0 \quad \begin{cases} \lambda_1 \leqslant 0,\ \lambda_2 \leqslant 0. \\ \lambda_1 \cdot \lambda_2 \leqslant 0 \quad (+,-) \text{ saddle pt.} \end{cases}$



$f(x,y) = x^2 - y^2$.

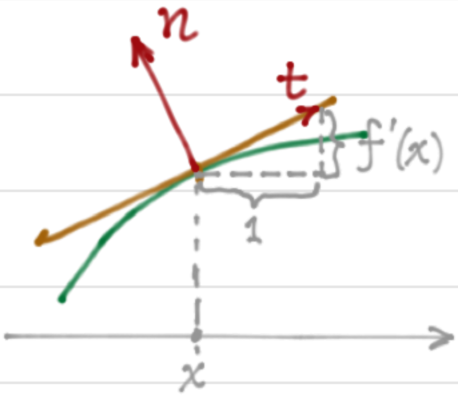
red pt: $(0,0)$

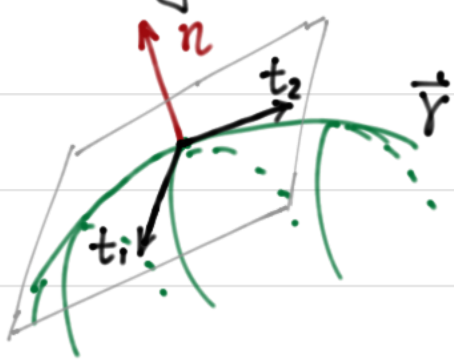
Assignment: Read Ch3 Transformations

Notation. $f: \mathbb{R}^n \to \mathbb{R}$ $\quad \vec{x} \in \mathbb{R}^n$, $\vec{x} = (x_1, x_2, \cdots, x_n)$

$Df \triangleq \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}\right)$, $\quad \nabla f = (Df)^T = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$

1-d, $f'(x)$ is the slope of the tagent line at x.





$\vec{r}: (x,y) \longrightarrow (x, y, f(x,y))$

a parametrization

$\vec{n} = (-f'(x), 1)$

$\vec{t} = (1, f'(x))$

$\vec{t}_1 = \frac{\partial \vec{r}}{\partial x} = (1, 0, \partial_x f)$

$\vec{t}_2 = \frac{\partial \vec{r}}{\partial y} = (0, 1, \partial_y f)$

$\vec{n} = \vec{t}_1 \times \vec{t}_2 = (-\partial_x f, -\partial_y f, 1)$

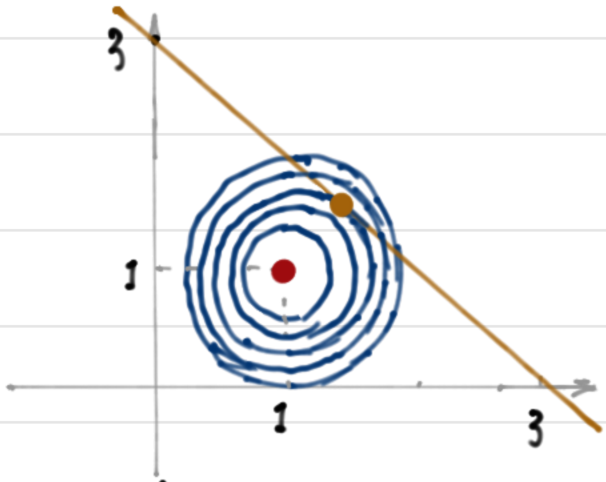
$H = \nabla^2 f = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}\right)$. As $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$, $\nabla^2 f$ is symmetric

Assignment: Read Ch5 Elements of Calculus

3 $\min f(x,y)$ s.t. $x+y=3$.

$f(x,y) = (x-1)^2 + (y-1)^2$, $\Omega = \{(x,y) \in \mathbb{R}^2 \mid x+y=3\}$

This is a 2-D, smooth, convex, and constrained optimization.



Level set $S_c = \{\vec{x} : f(\vec{x}) = c\}$

Level set S_c is a curve in $\mathbb{R}^2$

Graph of f $\{(\vec{x}, f(\vec{x})) \in \mathbb{R}^3\}$ is a surface in $\mathbb{R}^3$.

non-constrained minimum is at $(1,1)$ (red pt) $\nabla f(1,1) = 0$.

With an equality constraint, the minimum is changed to brown where $\nabla f \neq 0$! More complicated optimality condition.

For this problem, we can eliminate y to get a 1-d non-constrained smooth, and convex opt problem.

$y = 3-x$ $\quad \tilde{f}(x) \triangleq f(x, 3-x) = (x-1)^2 + (x-2)^2$.

$\tilde{f}'(x) = 0$. so $x = \frac{3}{2}$, $y = \frac{3}{2}$. $\quad \tilde{f}''(x) = 4 > 0 \ \forall x$

$x = \frac{3}{2}$ is a local minimum and as $\tilde{f}$ is convex, it is a global min.

so $\min f(x,y)$ s.t. $x+y=3$ is $\frac{1}{2}$ and the minimum pt $(\frac{3}{2}, \frac{3}{2})$

Fact: for a convex function, a local minimum is also a global one (to be proved soon)

Level sets and gradient.

$S_c = \{ \vec{x} \mid f(\vec{x}) = c \}$. This is a smooth curve for most c.

How to represent / describe a curve? Parametrization.

$g: \mathbb{R} \to \mathbb{R}^n$. $g(t) = (x(t), y(t))$ in $\mathbb{R}^2$.

$h: \mathbb{R} \to \mathbb{R}$ $\quad h(t) \triangleq f(g(t))$

$$\begin{tikzcd} & \mathbb{R}^n \arrow[dr, "f"] & \\ \mathbb{R} \arrow[ur, "g"] \arrow[rr, "h"] & & \mathbb{R} \end{tikzcd}$$

$h(t) = c$ by definition. So $h'(t) = 0$.

By chain rule, $h'(t) = \nabla f(g(t)) \cdot g'(t)$. So for a pt $\vec{x}_0 \in S_c$, we have $\nabla f(x_0) \cdot \vec{v} = 0$ where $\vec{v}$ is a tagent vector of S_c at $\vec{x}_0$

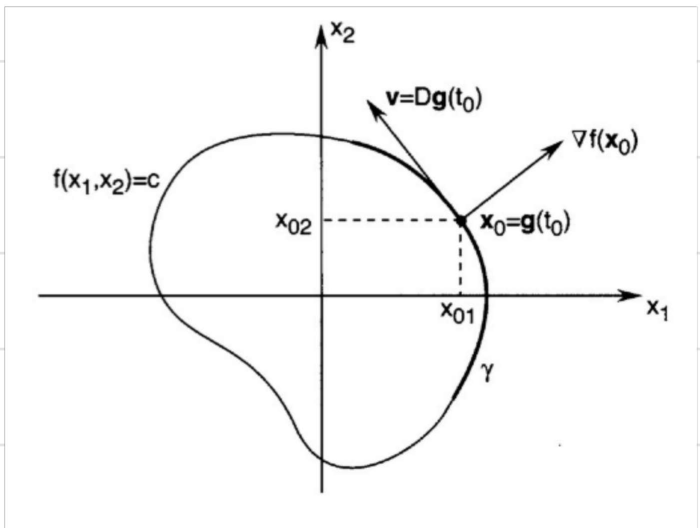


Figure 5.4 Orthogonality of the gradient to the level set.

Theorem. $\nabla f(x_0) \perp \vec{v}$,

$\forall \vec{v}$ tangent vector at $\vec{x}_0$ of the level set S_c for $c = f(x_0)$

§5.5 Page 68–72.

$\nabla f(x)$ is the direction of maximum rate of increase of f at x.

$-\nabla f(x)$ is the direction of maximum rate of decrease of f at x.

$-\nabla f(x)$: steepest descent direction

4 Rosenbrock function $f(x,y) = 100(y-x^2)^2 + (1-x)^2$

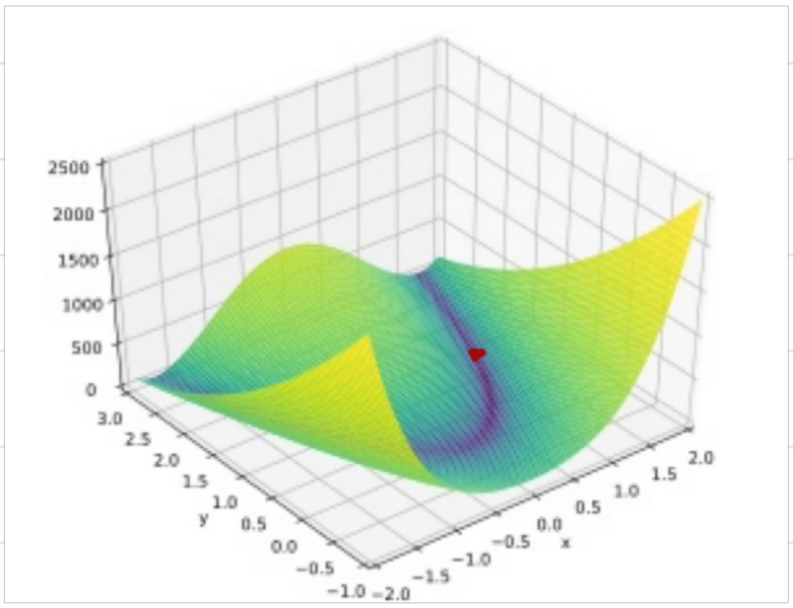
$\min_{\vec{x} \in \mathbb{R}^2} f(\vec{x})$ non-constrained, smooth, but non-convex

$$\nabla f(\vec{x}) = \begin{pmatrix} -400\,x_1(x_2 - x_1^2) - 2(1-x_1) \\ 200(x_2 - x_1^2) \end{pmatrix}, \quad \text{critical point } (1,1)$$

$$H(\vec{x}) = \nabla^2 f(\vec{x}) = \begin{pmatrix} 1200\,x_1^2 - 400\,x_2 + 2 & -400\,x_1 \\ -400\,x_1 & 200 \end{pmatrix}$$

$$H(1,1) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix} > 0.$$ So $(1,1)$ is a local minimum

Can show $(1,1)$ is a global minimum. It is inside a long, narrow, parabolic shaped flat valley.



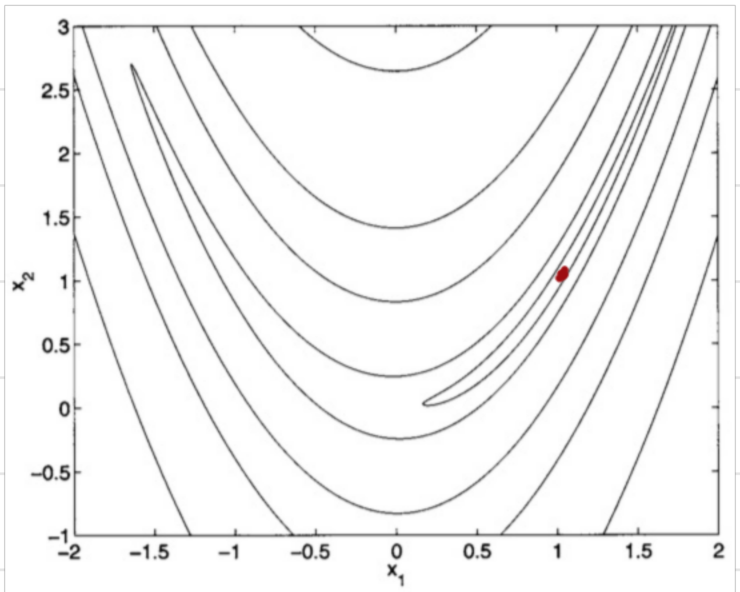


Figure 5.3 Level sets of Rosenbrock's (banana) function.