Nonlinear Optimization

Definitions

Niclas Börlin

Department of Computing Science

Umeå University
niclas.borlin@cs.umu.se

November 8, 2007

The continuous optimization problem

- In its most general form, the continuous optimization problems we will study may be written

$$\min_{x\in\Re^n} f(x) \qquad \text{subject to} \qquad \begin{array}{ll} c_i(x) = 0, & i \in \mathcal{E} \\ c_i(x) \geq 0, & i \in \mathcal{I} \end{array}$$

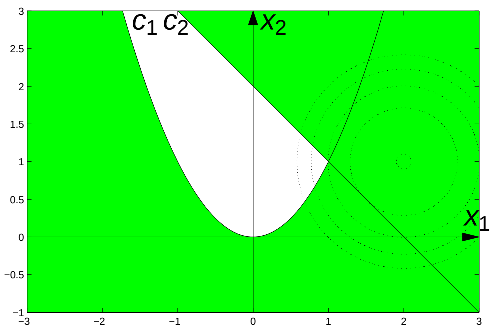
- The function $f(x) : \Re^n \to \Re$ is called the *objective function* and is assumed to be twice continuously differentiable.
- The vector x contains the *variables* to be estimated.
- The functions $c_i(x) : \Re^n \to \Re$ define *constraints* on the variables.
- The sets $\mathcal{E}$ and $\mathcal{I}$ are index sets for the equality and inequality constraints, respectively.
- A *maximization* problem is rewritten as

$$\max_x f(x) \equiv -\min_x -f(x).$$

Example

- Consider the problem

$$\min (x_1-2)^2 + (x_2-1)^2$$
$$\text{subject to} \quad x_1^2 - x_2 \leq 0,$$
$$x_1 + x_2 \leq 2.$$

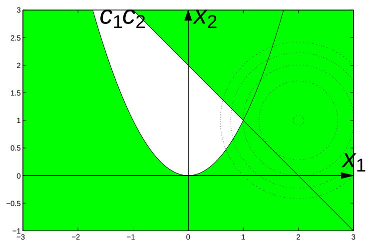


- We may rewrite this problem into general form by defining

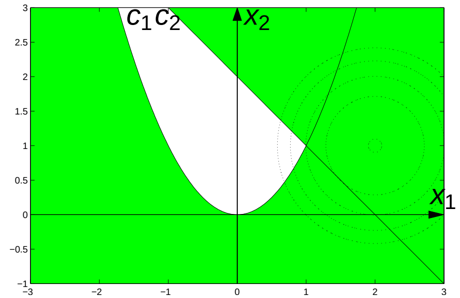
$$f(x) = (x_1-2)^2 + (x_2-1)^2,$$
$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$
$$c(x) = \begin{bmatrix} c_1(x) \\ c_2(x) \end{bmatrix} = \begin{bmatrix} -x_1^2 + x_2 \\ -x_1 - x_2 + 2 \end{bmatrix},$$
$$\mathcal{I} = \{1, 2\}, \mathcal{E} = \emptyset.$$

The parameter space

- The vector x will be interpreted as a point in $\Re^n$, the parameter space.
- Points that satisfies all constraints are called *feasible* and belong to the feasible set Ω which is a subset of $\Re^n$.
- At a feasible point x, an inequality constraint $c_i(x) \geq 0$ is said to be *binding* or *active* if $c_i(x) = 0$.
- If $c_i(x) > 0$, the constraint is *nonbinding* or *inactive*.
- Equality constraints are always active.



- The point x is said to be on the *boundary* of the constraint if $c_i(x) = 0$ and in the *interior* of the constraint if $c_i(x) > 0$.
- Equality constraints have no interior points.
- The set of active constraints at a given point is called the *active set* (of constraints).
- A feasible point with at least one active constraint belongs to the boundary of the feasible set.
- All other points are interior points to the feasible set.



Convexity

- If the objective function and the feasible set are *convex*, then the problem is much easier to solve.
- A set S is *convex* if

$$\alpha x + (1-\alpha)y \in S,\ 0 \le \alpha \le 1,\ \ \forall (x, y) \in S,$$

  i.e. all lines between all point-pairs in S is within S.
- A function f is convex on a convex set S if

$$f(\alpha x + (1-\alpha)y) \le \alpha f(x) + (1-\alpha)f(y),\ 0 \le \alpha \le 1,\ \forall x, y \in S,$$

  i.e. the function f is on or below the line through $(x, f(x))$ and $(y, f(y))$.

Global optimizers

- Consider the n-dimensional problem problemet

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

- A point x^* that satisfies

$$f(x^*) \le f(x)\ \forall x$$

  is called a *global minimizer* to f.
- The point x^* is often called a *solution* or *solution point*.
- If

$$f(x^*) < f(x)\ \forall x \ne x^*,$$

  the point x^* is called a *strict global minimizer* and is unique.
- Global minimizer are hard to determine unless f have special properties (e.g. is convex).

Local optimizers

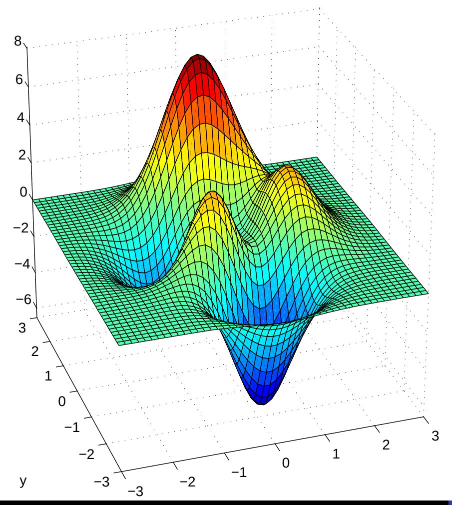
- Often we will have to settle for *local minimizers*, i.e. x^* such that

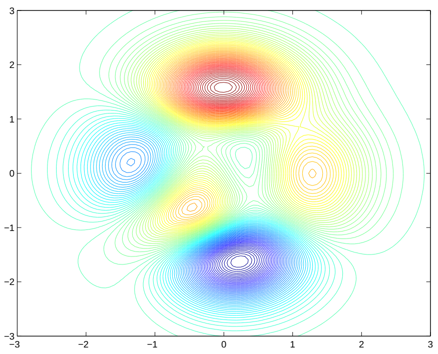
$$f(x^*) \le f(x)\ \forall x \in \mathcal{N},$$

  where $\mathcal{N}$ is a *neighbourhood* of x^*, i.e. an open set that contains x^*.
- Similarly, a *strict local minimizer* x^* is defined by

$$f(x^*) < f(x)\ \forall x \in \mathcal{N}, x \ne x^*.$$

Example





Taylor series

- The Taylor series is a tool for approximating a function f near a specific point x_0.
- The Taylor series may be applied whenever the function has derivatives and has many uses:
  - It enables approximation of a function value near a given point if the function itself is difficult to evaluate.
  - The approximation polynomial is easy to differentiate and integrate.
  - It is used to derive many algorithms for finding zeroes of function, for minimizing function, etc.
- Definition: Let x_0 be a specified point and $f : \Re \to \Re$ have n continous derivatives. Then the n-th order Taylor series approximation is

$$f(x_0 + p) \approx f(x_0) + pf'(x_0) + \frac{p^2}{2} f''(x_0) + \cdots + \frac{p^n}{n!} f^{(n)}(x_0).$$

Multi-dimensional Taylor series

- It is possible to defined Taylor series for real-values function of more than one variable:

$$1 \text{ variable } (x_0, p \in \Re)\ f(x_0 + p) = f(x_0) + pf'(x_0) + \frac{1}{2} pf''(x_0)p + \cdots$$

$$n \text{ variables } (x_0, p \in \Re^n)\ f(x_0 + p) = f(x_0) + p^T \nabla f(x_0) + \frac{1}{2} p^T \nabla^2 f(x_0) p + \cdots$$

- The notation $\nabla f(x_0)$ refers to the *gradient* of the function f at the point $x = x_0$, i.e. a column vector with all first order derivatives of f as elements $\frac{\partial f}{\partial x_i}(x_0)$.
- The notation $\nabla^2 f(x_0)$ refers to the *hessian* of f at x_0, i.e. a square matrix with all second order derivatives of f as elements $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_0)$.

Example

- For the function $f(x_1, x_2) = x_1^3 + 5x_1^2 x_2 + 7x_1 x_2^2 + 2x_2^3$, the gradient and hessian are

$$\nabla f(x) = \begin{bmatrix} 3x_1^2 + 10x_1x_2 + 7x_2^2 \\ 5x_1^2 + 14x_1x_2 + 6x_2^2 \end{bmatrix}, \quad \nabla^2 f(x) = \begin{bmatrix} 6x_1 + 10x_2 & 10x_1 + 14x_2 \\ 10x_1 + 14x_2 & 14x_1 + 12x_2 \end{bmatrix}.$$

- Evaluated in the point $x_0 = [-2, 3]^T$ they are

$$\nabla f(x_0) = \begin{bmatrix} 15 \\ -10 \end{bmatrix} \text{ and } \nabla^2 f(x_0) = \begin{bmatrix} 18 & 22 \\ 22 & 8 \end{bmatrix}.$$

- For $p = [0.1, 0.2]^T$,

$$\begin{aligned} f(-1.9, 3.2) = f(x_0 + p) &\approx f(x_0) + p^T \nabla f(x_0) + \frac{1}{2} p^T \nabla^2 f(x_0) p \\ &= -20 + [0.1\ 0.2] \begin{bmatrix} 15 \\ -10 \end{bmatrix} + [0.1\ 0.2] \begin{bmatrix} 18 & 22 \\ 22 & 8 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.2 \end{bmatrix} \\ &= -20 - 0.5 + 0.69 = -19.81 \end{aligned}$$

- Compare the exact value $f(-1.9, 3.2) = -19.755$.

Definiteness

- A square matrix A is *positive semi-definite* if

$$x^T A x \geq 0, \forall x.$$

- A square matrix A is *positive definite* if

$$x^T A x > 0, \forall x \neq 0.$$

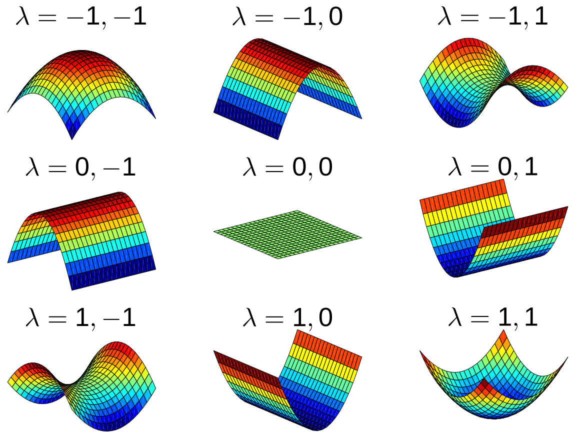
- A square matrix A which is neither positive semi-definite nor negative semi-definite is *indefinite*.
- A positive definite matrix has only positive eigenvalues since for all eigenpairs (x, λ) of A

$$x^T A x > 0 \Rightarrow x^T \underbrace{Ax}_{\lambda x} = x^T \lambda x = \lambda x^T x > 0 \Rightarrow \lambda > 0.$$

- Positive semi-definite matrices have $\lambda \geq 0$.

Definiteness and curvature

- The definiteness and sign of the eigenvalues correspond to *curvatures* of the quadratic expression $x^T A x$.



First-Order Necessary Conditions

- Assume x^* is a local minimizer to f. Study f around x^*:

$$f(x^* + p) = f(x^*) + \nabla f(x^*)^T p + \frac{1}{2} p^T \nabla^2 f(\xi) p$$

- If x^* is a local minimizer this implies that

$$\nabla f(x^*) = 0.$$

- This condition is called the $\boxed{\text{first-order } \textit{necessary} \text{ condition}}$ for a minimizer.
- A point that satisfies $\nabla f(x^*) = 0$ is called a *stationary point* to f.

Second-Order Necessary Conditions

- Study f around a stationary point x^*. Assume x^* is a local minimizer:

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\nabla f(x^*)^T p}_{=0} + \frac{1}{2} p^T \nabla^2 f(\xi) p.$$

- For x close to x^*, $\nabla^2 f(\xi)$ will be close to $\nabla^2 f(x^*)$.
- If $\nabla^2 f(x^*)$ is *not* positive semi-definite, there exists a v such that $v^T \nabla^2 f(x^*) v < 0$, and there is a p close to v such that $p^T \nabla^2 f(\xi) p < 0$ meaning that

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\frac{1}{2} p^T \nabla^2 f(\xi) p}_{<0} < f(x^*),$$

which is a contradiction, since x^* was assumed to be a minimizer.

Second-Order Necessary Conditions

- Thus, $\nabla^2 f(x^*)$ must be positive semi-definite in order for the stationary point x^* to be a minimizer.
- This condition is called the $\boxed{\text{second-order } \textit{necessary} \text{ condition}}$.

Second-Order Sufficient Conditions

- Study f around x^* when $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*)$ positive definite.
- Then

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\nabla f(x^*)^T p}_{=0} + \frac{1}{2} p^T \nabla^2 f(\xi) p.$$

- For x close to x^*, $\nabla^2 f(\xi)$ will also be positive definite and

$$f(x) = f(x^* + p) = f(x^*) + \underbrace{\frac{1}{2} p^T \nabla^2 f(\xi) p}_{>0 \ \forall p \neq 0} > f(x^*) \quad \forall p$$

- Thus, x^* is a strict minimizer of f.
- This is called the $\boxed{\text{second-order } \textit{sufficient} \text{ condition}}$ on a minimizer.