Convexity with respect to generalized inequalities

- Monotonicity with respect to a generalized inequality

Suppose $K \subseteq \mathbb{R}^n$ is a proper cone with associated generalized inequality $\preceq_K$. A function $f : \mathbb{R}^n \to \mathbb{R}$ is called K-*nondecreasing* if

$$\mathbf{x} \preceq_K \mathbf{y} \Rightarrow f(\mathbf{x}) \leq f(\mathbf{y})$$

and K-*increasing* if

$$\mathbf{x} \preceq_K \mathbf{y}, \mathbf{x} \neq \mathbf{y} \Rightarrow f(\mathbf{x}) < f(\mathbf{y}).$$

We define K-*nonincreasing* and K-*decreasing* functions in a similar way.

Convexity with respect to generalized inequalities

- Monotonicity with respect to a generalized inequality

Examples. 1. Monotone vector functions. A function $f : \mathbb{R}^n \to \mathbb{R}$ is nondecreasing with respect to $\mathbb{R}^n_+$ iff

$$x_1 \leq y_1, ..., x_n \leq y_n \Rightarrow f(\mathbf{x}) \leq f(\mathbf{y})$$

for all $\mathbf{x}, \mathbf{y}$. This is the same as saying that f, when restricted to any component x_i (*i.e.*, x_i is considered the variable while x_j for $j \neq i$ are fixed), is nondecreasing.

Convexity with respect to generalized inequalities

- Monotonicity with respect to a generalized inequality

Examples. 2. Matrix monotone functions. A function $f : \mathbb{S}^n \to \mathbb{R}$ is called *matrix monotone* (increasing, decreasing) if it is monotone with respect to the positive semidefinite cone. Some examples of matrix monotone functions of the variable $\mathbf{X} \in \mathbb{S}^n$:

- $\mathrm{tr}(\mathbf{WX})$,where $\mathbf{W} \in \mathbb{S}^n$, is matrix nondecreasing if $\mathbf{W} \succeq \mathbf{0}$, and matrix increasing if $\mathbf{W} \succ \mathbf{0}$ (it is matrix nonincreasing if $\mathbf{W} \preceq \mathbf{0}$, and matrix decreasing if $\mathbf{W} \prec \mathbf{0}$).
- $\mathrm{tr}(\mathbf{X}^{-1})$ is matrix decreasing on $\mathbb{S}^n_{++}$.
- $\det \mathbf{X}$ is matrix increasing on $\mathbb{S}^n_{++}$.

Convexity with respect to generalized inequalities

- Gradient conditions for monotonicity

A differentiable function f, with convex domain, is K-nondecreasing if and only if

$$\nabla f(\mathbf{x}) \succeq_{K^*} \mathbf{0} \tag{1}$$

for all $\mathbf{x} \in \mathbf{dom}\, f$. Note the difference with the simple scalar case: the gradient must be nonnegative in the *dual* inequality.

For the strict case, we have the following: If

$$\nabla f(\mathbf{x}) \succ_{K^*} \mathbf{0} \tag{2}$$

for all $\mathbf{x} \in \mathbf{dom}\, f$, then f is K-increasing. As in the scalar case, the converse is not true.

Convexity with respect to generalized inequalities

- Convexity with respect to a generalized inequality

We say $f : \mathbb{R}^n \to \mathbb{R}^m$ is *K-convex* if for all $\mathbf{x}, \mathbf{y}$, and $0 \leq \theta \leq 1$,

$$f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \preceq_K \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}).$$

The function is *strictly K-convex* if

$$f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \prec_K \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})$$

for all $\mathbf{x} \neq \mathbf{y}$ and $0 < \theta < 1$. These definitions reduce to ordinary convexity and strict convexity when $m = 1$ (and $K = \mathbb{R}_+$).

Convexity with respect to generalized inequalities

- Examples

1. Convexity with respect to componentwise inequality. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is convex with respect to componentwise inequality (*i.e.*, the generalized inequality induced by $\mathbb{R}^m_+$) iff for all $\mathbf{x}, \mathbf{y}$ and $0 \leq \theta \leq 1$,

$$f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \preceq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}),$$

i.e., each component f_i is a convex function. The function f is strictly convex with respect to componentwise inequality iff each component f_i is strictly convex.

Convexity with respect to generalized inequalities

- Examples

2. Matrix convexity. $f : \mathbb{R}^n \to \mathbb{S}^m$ is convex w.r.t. matrix inequality if

$$f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \preceq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})$$

for any $\mathbf{x}$ and $\mathbf{y}$, and for $\theta \in [0,1]$. This is sometimes called *matrix convexity*. **An equivalent definition is that the scalar function $\mathbf{z}^T f(\mathbf{x})\mathbf{z}$ is convex for all vectors $\mathbf{z}$.** A matrix function is strictly matrix convex if

$$f(\theta \mathbf{x} + (1-\theta)\mathbf{y}) \prec \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y})$$

when $\mathbf{x} \neq \mathbf{y}$ and $0 < \theta < 1$, or, equivalently, if **$\mathbf{z}^T f(\mathbf{x})\mathbf{z}$ is strictly convex for every $\mathbf{z} \neq \mathbf{0}$.** Some examples:

- The function $f(\mathbf{X}) = \mathbf{X}\mathbf{X}^T$ where $\mathbf{X} \in \mathbb{R}^{n \times m}$ is matrix convex.
- The function $\mathbf{X}^p$ is matrix convex on $\mathbb{S}^n_{++}$ for $1 \leq p \leq 2$ or $-1 \leq p \leq 0$, and matrix concave for $0 \leq p \leq 1$.
- The function $f(\mathbf{X}) = e^{\mathbf{X}}$ is not matrix convex on $\mathbb{S}^n$, for $n \geq 2$.

Convexity with respect to generalized inequalities

- Examples

Many of the results for convex functions have extensions to K-convex functions, e.g., a function is K-convex if and only if its restriction to any line in its domain is K-convex.

Dual characterization of K-convexity

A function f is K-convex iff for every $\boldsymbol{\omega} \succeq_{K^*} \mathbf{0}$, the (real-valued) function $\boldsymbol{\omega}^T f$ is convex (in the ordinary sense); f is strictly K-convex iff for every nonzero $\boldsymbol{\omega} \succeq_{K^*} \mathbf{0}$ the function $\boldsymbol{\omega}^T f$ is strictly convex.

Differentiable K-convex functions

A differentiable function f is K-convex iff its domain is convex, and for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom}\, f$,

$$f(\mathbf{y}) \succeq_K f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{y} - \mathbf{x}).$$

Here $Df(\mathbf{x}) \in \mathbb{R}^{m \times n}$ is the derivative of f at $\mathbf{x}$. The function f is strictly K-convex iff for all $\mathbf{x}, \mathbf{y} \in \mathbf{dom}\, f$ with $\mathbf{x} \neq \mathbf{y}$,

$$f(\mathbf{y}) \succ_K f(\mathbf{x}) + Df(\mathbf{x})(\mathbf{y} - \mathbf{x}).$$

Convexity with respect to generalized inequalities

- Examples

Composition theorem

Many of the results on composition can be generalized to K-convexity. For example, if $g : \mathbb{R}^n \to \mathbb{R}^p$ is K-convex, $h : \mathbb{R}^p \to \mathbb{R}$ is convex, and $\tilde{h}$ (the extended-value extension of h) is K-nondecreasing, then $h \circ g$ is convex. This generalizes the fact that a nondecreasing convex function of a convex function is convex. The condition that $\tilde{h}$ be K-nondecreasing implies that $\mathbf{dom}\, h - K = \mathbf{dom}\, h$.

Convexity with respect to generalized inequalities

- Examples

Example. The quadratic matrix function $g : \mathbb{R}^{m\times n} \to \mathbb{S}^n$ defined by

$$g(\mathbf{X}) = \mathbf{X}^T\mathbf{A}\mathbf{X} + \mathbf{B}^T\mathbf{X} + \mathbf{X}^T\mathbf{B} + \mathbf{C},$$

where $\mathbf{A} \in \mathbb{S}^m$, $\mathbf{B} \in \mathbb{R}^{m\times n}$, and $\mathbf{C} \in \mathbb{S}^n$, is convex when $\mathbf{A} \succeq \mathbf{0}$.

The function $h : \mathbb{S}^n \to \mathbb{R}$ defined by $h(\mathbf{Y}) = -\log\det(-\mathbf{Y})$ is convex and increasing on $\mathbf{dom}\, h = -\mathbb{S}^n_{++}$. By the composition theorem, we conclude that

$$f(\mathbf{X}) = -\log\det(-(\mathbf{X}^T\mathbf{A}\mathbf{X} + \mathbf{B}^T\mathbf{X} + \mathbf{X}^T\mathbf{B} + \mathbf{C}))$$

is convex on $\mathbf{dom}\, f = \{\mathbf{X} \in \mathbb{R}^{m\times n} \mid \mathbf{X}^T\mathbf{A}\mathbf{X} + \mathbf{B}^T\mathbf{X} + \mathbf{X}^T\mathbf{B} + \mathbf{C} \prec \mathbf{0}\}$.

This generalizes the fact that $-\log(-(ax^2 + bx + c))$ is convex on $\{x \in \mathbb{R} \mid ax^2 + bx + c < 0\}$, provided $a \geq 0$.

A little about nonconvex analysis

- Basic concepts

Definition 1. *A function $g : \mathbb{R}^n \to (-\infty, +\infty]$ is said to be lower semicontinuous at point $\mathbf{x}_0$ if*

$$\liminf_{x \to \mathbf{x}_0} g(\mathbf{x}) \geq g(\mathbf{x}_0). \tag{1}$$

Definition 2. *A function $F(\mathbf{x})$ is called coercive if F is bounded from below and*

$$F(\mathbf{x}) \to \infty \quad \textit{when} \quad \|\mathbf{x}\| \to \infty, \tag{2}$$

where $\|\cdot\|$ is the l_2-norm.

A little about nonconvex analysis

- Basic concepts

Definition 3. *Let g be a proper and lower semicontinuous function.*

1. *For a given* $\mathbf{x} \in$*dom* g, *the Frechet subdifferential of* g *at* $\mathbf{x}$, *written as* $\hat{\partial} g(\mathbf{x})$, *is the set of all vectors* $\mathbf{u} \in \mathbb{R}^n$ *which satisfy*

$$\liminf_{\mathbf{y} \neq \mathbf{x}, \mathbf{y} \to \mathbf{x}} \frac{g(\mathbf{y}) - g(\mathbf{x}) - \langle \mathbf{u}, \mathbf{y} - \mathbf{x} \rangle}{\|\mathbf{y} - \mathbf{x}\|} \geq 0. \tag{1}$$

2. *The limiting-subdifferential, or simply the subdifferential, of* g *at* $\mathbf{x} \in \mathbb{R}^n$, *written as* $\partial g(\mathbf{x})$, *is defined through the following closure process*

$$\partial f(\mathbf{x}) := \{\mathbf{u} \in \mathbb{R}^n : \exists \mathbf{x}_k \to \mathbf{x}, g(\mathbf{x}_k) \to g(\mathbf{x}), \mathbf{u}_k \in \hat{\partial} g(\mathbf{x}_k) \to \mathbf{u}, k \to \infty\}. \tag{2}$$

A little about nonconvex analysis

- Basic concepts

Proposition 1.

1. *In the nonsmooth context, the Fermat's rule remains unchanged: If* $\mathbf{x} \in \mathbb{R}^n$ *is a local minimizer of* g, *then* $\mathbf{0} \in \partial g(\mathbf{x})$.

2. *Let* $(\mathbf{x}_k, \mathbf{u}_k)$ *be a sequence such that* $\mathbf{x}_k \to \mathbf{x}$, $\mathbf{u}_k \to \mathbf{u}$, $g(\mathbf{x}_k) \to g(\mathbf{x})$ *and* $\mathbf{u}_k \in \partial g(\mathbf{x}_k)$, *then* $\mathbf{u} \in \partial g(\mathbf{x})$.

3. *If* f *is a continuously differentiable function, then* $\partial(f+g)(\mathbf{x}) = \nabla f(\mathbf{x}) + \partial g(\mathbf{x})$.

Points whose subdifferential contains $\mathbf{0}$ are called critical points.

Chapter 5. Unconstrained Optimization

- Unconstrained minimization problems
- Descent methods
- Newton’s method
- Conjugate direction methods
- Quasi-Newton methods
- Majorization minimization

Unconstrained minimization problems

- Model problem

$$\min_{\mathbf{x}} \quad f(\mathbf{x}), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is convex and twice continuously differentiable (which implies that $\mathbf{dom} f$ is open). We will assume that the problem is solvable, *i.e.*, there exists an optimal point $\mathbf{x}^*$.

Since f is differentiable and convex, a necessary and sufficient condition for a point $\mathbf{x}^*$ to be optimal is

Optimility condition

$$\boxed{\nabla f(\mathbf{x}^*) = \mathbf{0}.} \tag{2}$$

Usually we need to design an algorithm that computes a sequence of points $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \ldots \in \mathbf{dom} f$ with $f(\mathbf{x}^{(k)}) \to p^*$ as $k \to \infty$. Such a sequence of points is called a minimizing sequence for the problem (1). The algorithm is terminated when $\boxed{f(\mathbf{x}^{(k)}) - p^* \le \epsilon,}$ where $\epsilon > 0$ is some specified tolerance.

Checked by: $\|\nabla f(\mathbf{x}^k)\| \le \varepsilon$ (stopping criterion)

Unconstrained minimization problems

- Initial point and sublevel set

The starting point $\mathbf{x}^{(0)}$ must lie in $\mathbf{dom}\, f$, and in addition the sublevel set

$$S = \{\mathbf{x} \in \mathbf{dom}\, f | f(\mathbf{x}) \le f(\mathbf{x}^{(0)})\}$$

must be closed.

Examples. 1. Quadratic minimization and least-squares:

$$\min_{\mathbf{x}} (1/2)\mathbf{x}^T \mathbf{P} \mathbf{x} + \mathbf{q}^T \mathbf{x} + r,$$

where $\mathbf{P} \in \mathbb{S}^n_+, \mathbf{q} \in \mathbb{R}^n$, and $r \in \mathbb{R}$.
2. Unconstrained geometric programming:

$$\min_{\mathbf{x}} \; f(\mathbf{x}) = \log \sum_{i=1}^{m} \exp(\mathbf{a}_i^T \mathbf{x} + b_i).$$

Unconstrained minimization problems

- Initial point and sublevel set

3. Analytic center of linear inequalities:

$$\min_{\mathbf{x}} \ f(\mathbf{x}) = -\sum_{i=1}^{m} \log(b_i - \mathbf{a}_i^T \mathbf{x}),$$

where $\mathbf{dom} f = \{\mathbf{x} | \mathbf{a}_i^T \mathbf{x} < b_i, i = 1, ..., m\}$. The objective function f in this problem is called the *logarithmic barrier* for the inequalities $\mathbf{a}_i^T \mathbf{x} \leq b_i$.

4. Analytic center of a linear matrix inequality:

$$\min_{\mathbf{x}} \ f(\mathbf{x}) = \log\det F(\mathbf{x})^{-1},$$

where $F : \mathbb{R}^n \to \mathbb{S}^p$is affine, *i.e.*,

$$F(\mathbf{x}) = \mathbf{F}_0 + x_1 \mathbf{F}_1 + \cdots + x_n \mathbf{F}_n,$$

with $\mathbf{F}_i \in \mathbb{S}^p$. $\mathbf{dom} f = \{\mathbf{x} | F(\mathbf{x}) \succ \mathbf{0}\}$. The objective function f is called the *logarithmic barrier* for the linear matrix inequality $F(\mathbf{x}) \succ \mathbf{0}$.

Unconstrained minimization problems

- Strong convexity and implications

Strong convexity:

$$\nabla^2 f(\mathbf{x}) \succeq m\mathbf{I}.$$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{m}{2}\|\mathbf{y}-\mathbf{x}\|_2^2, \tag{1}$$

Bound $f(\mathbf{x}) - p^*$ in terms of $\|\nabla f(\mathbf{x})\|_2$. The righthand side of (1) is a convex quadratic function of $\mathbf{y}$ (for fixed $\mathbf{x}$). Setting the gradient with respect to $\mathbf{y}$ equal to zero, we find that $\tilde{\mathbf{y}} = \mathbf{x} - (1/m)\nabla f(\mathbf{x})$ minimizes the righthand side. Therefore we have

$$\begin{aligned} f(\mathbf{y}) &\geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{m}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 \\ &\geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\tilde{\mathbf{y}}-\mathbf{x}) + \frac{m}{2}\|\tilde{\mathbf{y}}-\mathbf{x}\|_2^2 \\ &= f(\mathbf{x}) - \frac{1}{2m}\|\nabla f(\mathbf{x})\|_2^2. \end{aligned}$$

Unconstrained minimization problems

- Strong convexity and implications

Since this holds for any $\mathbf{y} \in S$, we have

$$p^* \geq f(\mathbf{x}) - \frac{1}{2m} \|\nabla f(\mathbf{x})\|_2^2. \tag{2}$$

$$\|\nabla f(\mathbf{x})\|_2 \leq (2m\epsilon)^{1/2} \Rightarrow f(\mathbf{x}) - p^* \leq \epsilon \tag{3}$$

Unconstrained minimization problems

- Strong convexity and implications

We can also derive a bound on $\|\mathbf{x}-\mathbf{x}^*\|_2$ in terms of $\|\nabla f(\mathbf{x})\|_2$.

$$\|\mathbf{x}-\mathbf{x}^*\|_2 \le \frac{2}{m}\|\nabla f(\mathbf{x})\|_2. \tag{4}$$

Proof. We apply (1) with $\mathbf{y}=\mathbf{x}^*$ to obtain

$$\begin{aligned} p^* = f(\mathbf{x}^*) &\ge f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{x}^*-\mathbf{x}) + \frac{m}{2}\|\mathbf{x}^*-\mathbf{x}\|_2^2 \\ &\ge f(\mathbf{x}) - \|\nabla f(\mathbf{x})\|_2\|\mathbf{x}^*-\mathbf{x}\|_2 + \frac{m}{2}\|\mathbf{x}^*-\mathbf{x}\|_2^2. \end{aligned}$$

Since $p^* \le f(\mathbf{x})$, we must have

$$-\|\nabla f(\mathbf{x})\|_2\|\mathbf{x}^*-\mathbf{x}\|_2 + \frac{m}{2}\|\mathbf{x}^*-\mathbf{x}\|_2^2 \le 0,$$

from which (4) follows. One consequence of (4) is that $\mathbf{x}^*$ is unique.

Unconstrained minimization problems

- Strong convexity and implications

Upper bound on $\nabla^2 f(\mathbf{x})$

Therefore the maximum eigenvalue of $\nabla^2 f(\mathbf{x})$, which is a continuous function of $\mathbf{x}$ on S, is bounded above on S, *i.e.*, there exists a constant M such that

$$\nabla^2 f(\mathbf{x}) \preceq M\mathbf{I}, \tag{5}$$

for all $\mathbf{x} \in S$. This upper bound on the Hessian implies for any $\mathbf{x}, \mathbf{y} \in S$,

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{M}{2}\|\mathbf{y} - \mathbf{x}\|_2^2, \tag{6}$$

which is analogous to (1). Minimizing each side over $\mathbf{y}$ yields

$$p^* \leq f(\mathbf{x}) - \frac{1}{2M}\|\nabla f(\mathbf{x})\|_2^2, \tag{7}$$

the counterpart of (2).

Unconstrained minimization problems

- Strong convexity and implications

Condition number of sublevel sets

We define the *width* of a convex set $C \subseteq \mathbb{R}^n$, in the direction $\mathbf{q}$, where $\|\mathbf{q}\|_2 = 1$, as

$$W(C, \mathbf{q}) = \sup_{\mathbf{z} \in C} \mathbf{q}^T \mathbf{z} - \inf_{\mathbf{z} \in C} \mathbf{q}^T \mathbf{z}.$$

The *minimum width* and maximum width of C are given by

$$W_{\min} = \inf_{\|\mathbf{q}\|_2 = 1} W(C, \mathbf{q}), \quad W_{\max} = \sup_{\|\mathbf{q}\|_2 = 1} W(C, \mathbf{q}).$$

The *condition number* of the convex set C is defined as

$$\mathbf{cond}(C) = \frac{W_{\max}^2}{W_{\min}^2},$$

i.e., the square of the ratio of its maximum width to its minimum width. The condition number of C gives a measure of its *anisotropy* or *eccentricity*.

Unconstrained minimization problems

- Strong convexity and implications

Example. *Condition number of an ellipsoid.* Let $\mathcal{E}$ be the ellipsoid

$$\mathcal{E} = \{\mathbf{x} | (\mathbf{x} - \mathbf{x}_0)^T \mathbf{A}^{-1} (\mathbf{x} - \mathbf{x}_0) \leq 1\},$$

where $\mathbf{A} \in \mathbb{S}^n_{++}$. The width of $\mathcal{E}$ in the direction $\mathbf{q}$ is

$$\sup_{\mathbf{z}\in\mathcal{E}} \mathbf{q}^T\mathbf{z} - \inf_{\mathbf{z}\in\mathcal{E}} \mathbf{q}^T\mathbf{z} = (\|\mathbf{A}^{1/2}\mathbf{q}\|_2 + \mathbf{q}^T\mathbf{x}_0) - (-\|\mathbf{A}^{1/2}\mathbf{q}\|_2 + \mathbf{q}^T\mathbf{x}_0) = 2\|\mathbf{A}^{1/2}\mathbf{q}\|_2.$$

It follows that its minimum and maximum width are

$$W_{\min} = 2\lambda_{\min}(\mathbf{A})^{1/2}, \quad W_{\max} = 2\lambda_{\max}(\mathbf{A})^{1/2}.$$

and its condition number is

$$\mathbf{cond}(\mathcal{E}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})} = \kappa(\mathbf{A}).$$

Unconstrained minimization problems

- Strong convexity and implications

Suppose f satisfies $m\mathbf{I} \preceq \nabla^2 f(\mathbf{x}) \preceq M\mathbf{I}$ for all $\mathbf{x} \in S$. We will derive a bound on the condition number of the α-sublevel $C_\alpha = \{\mathbf{x} | f(\mathbf{x}) \leq \alpha\}$, where $p^* < \alpha \leq f(\mathbf{x}^{(0)})$. We have

$$p * + (M/2)\|\mathbf{y} - \mathbf{x}^*\|_2^2 \geq f(\mathbf{y}) \geq p^* + (m/2)\|\mathbf{y} - \mathbf{x}^*\|_2^2.$$

This implies that $B_{inner} \subseteq C_\alpha \subseteq B_{outer}$ where

$$\begin{aligned} B_{inner} =& \{\mathbf{y} | \|\mathbf{y} - \mathbf{x}^*\|_2 \leq (2(\alpha - p^*)/M)^{1/2}\}, \\ B_{outer} =& \{\mathbf{y} | \|\mathbf{y} - \mathbf{x}^*\|_2 \leq (2(\alpha - p^*)/m)^{1/2}\}. \end{aligned}$$

The ratio of the radii squared gives an upper bound on the condition number of C_α:

$$\mathbf{cond}(C_\alpha) \leq \frac{M}{m}.$$

Descent methods

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

search direction

step size

where $t^{(k)} > 0$ (except when $\mathbf{x}^{(k)}$ is optimal).

Descent methods mean that

$$f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)}),$$

except when $\mathbf{x}^{(k)}$ is optimal.

From convexity we know that $\nabla f(\mathbf{x}^{(k)})^T(\mathbf{y} - \mathbf{x}^{(k)}) \geq 0$ implies $f(\mathbf{y}) \geq f(\mathbf{x}^{(k)})$, so the search direction in a descent method must satisfy

$$\nabla f(\mathbf{x}^{(k)})^T \Delta \mathbf{x}^{(k)} < 0,$$

i.e., it must make an acute angle with the negative gradient. We call such a direction a *descent direction* (for f, at $\mathbf{x}^{(k)}$).

Descent methods

Algorithm 5.1 *General descent method.*

given a starting point $\mathbf{x} \in \mathbf{dom} f$.

repeat

1. Determine a descent direction $\Delta\mathbf{x}$.
2. Line search. Choose a step size $t > 0$.
3. Update. $\mathbf{x} := \mathbf{x} + t\Delta\mathbf{x}$.

until stopping criterion is satisfied.

The stopping criterion is often of the form $\|\nabla f(\mathbf{x})\|_2 \leq \eta$, where η is small and positive.

Descent methods

- Exact line search

$$t = \underset{s \geq 0}{\operatorname{argmin}} f(\mathbf{x} + s\Delta\mathbf{x}).$$

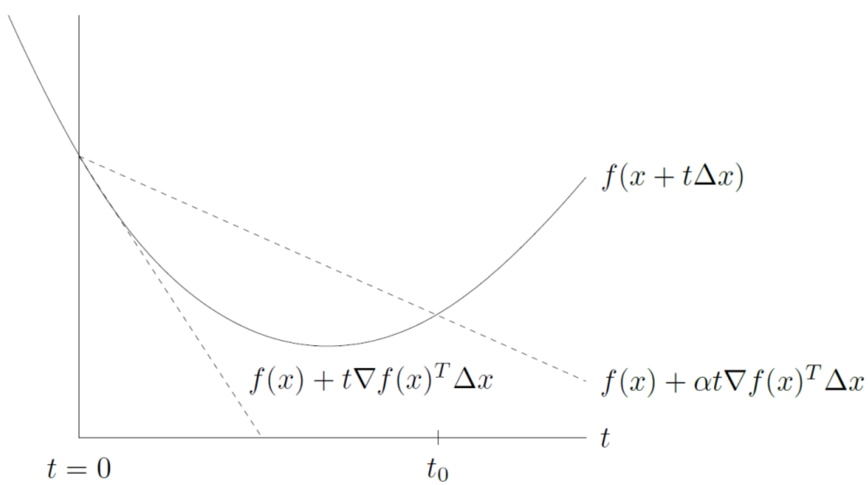
- Backtracking line search

Algorithm 9.2 *Backtracking line search.*

given a descent direction $\Delta\mathbf{x}$ for f at $\mathbf{x} \in \mathbf{dom} f$, $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$.

$t := 1.$

while $f(\mathbf{x} + t\Delta\mathbf{x}) > f(\mathbf{x}) + \alpha t \nabla f(\mathbf{x})^T \Delta\mathbf{x}$, $t := \beta t$.



Descent methods

- Backtracking line search

The line search is called *backtracking* because it starts with unit step size and then reduces it by the factor β until the stopping condition

$$f(\mathbf{x} + t\Delta\mathbf{x}) \leq f(\mathbf{x}) + \alpha t \nabla f(\mathbf{x})^T \Delta\mathbf{x}$$

holds. Since $\Delta\mathbf{x}$ is a descent direction, we have $\nabla f(\mathbf{x})^T \Delta\mathbf{x} < 0$, so for small enough t we have

$$f(\mathbf{x} + t\Delta\mathbf{x}) \;\approx f(\mathbf{x}) + t\nabla f(\mathbf{x})^T \Delta\mathbf{x} < f(\mathbf{x}) + \alpha t \nabla f(\mathbf{x})^T \Delta\mathbf{x},$$

which shows that the backtracking line search eventually terminates.

Descent methods

- Gradient descent method

A natural choice for the search direction is the negative gradient

$$\Delta \mathbf{x} = -\nabla f(\mathbf{x}).$$

The resulting algorithm is called the *gradient algorithm* or *gradient descent method*.

Algorithm 5.3 *Gradient descent method.*

given a starting point $\mathbf{x} \in \mathbf{dom} f$.

repeat

1. $\Delta \mathbf{x} := -\nabla f(\mathbf{x})$.
2. Line search. Choose step size t via exact or backtracking line search.
3. Update. $\mathbf{x} := \mathbf{x} + t\Delta \mathbf{x}$.

until stopping criterion is satisfied.

The stopping criterion is usually of the form $\|\nabla f(\mathbf{x})\|_2 \leq \eta$, where η is small and positive.