Unit: Optimality Conditions
(Analytical Solution of Multiobjective Optimization)
Goals

1. What is the Gradient of a function? What are its properties?
2. How can it be used to find a linear approximation of a nonlinear function?
3. Given a continuously differentiable function, which equations are fulfilled for local optima in the following cases?
   1. Unconstrained
   2. Equality Constraints
   3. Inequality Constraints
4. How can this be used to find Pareto fronts analytically?
5. How to state conditions for locally efficient in multiobjective optimization?
Optimality conditions for differentiable problems

- Given a point on a continuous differentiable function, often necessary (and sufficient) conditions can be stated for the point to be a local extremum.

\[ f'(x) = 0, \quad f''(x) < 0 \]

- *Necessary* conditions, can be used to restrict the set of candidate solutions.
- If *sufficient* conditions are met, this implies the solution is locally (Pareto) optimal, but we may have stated more than necessary.
Linear Taylor Approximations

Continuously differentiable functions can locally be approximated by tangent planes, i.e.

$$\lim_{x \to x_0} f(x) - [f(x_0) + \nabla f(x_0)(x - x_0)] = 0$$

$$\nabla f(x) = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})^T(x)$$
Gradient computation

\[ f(x) = 20 \exp(-(x_1)^2 - (x_2)^2) \]
\[ \frac{\partial f}{\partial x_1}(x) = 20 \cdot (-2x_1) \exp(-(x_1)^2 - (x_2)^2) \quad \text{chain rule} \]
\[ \frac{\partial f}{\partial x_2}(x) = 20 \cdot (-2x_2) \exp(-(x_1)^2 - (x_2)^2) \]
\[ \nabla f(x) = \begin{pmatrix} 20 \cdot (-2x_1) \exp(-(x_1)^2 - (x_2)^2) \\ 20 \cdot (-2x_2) \exp(-(x_1)^2 - (x_2)^2) \end{pmatrix} \]

Gradient vector at \((1, 1)\) is given by \((-40 \exp(-2), -40 \exp(-2))^\top\):

---

Program: wxMaxima:

```maxima
load(draw);
draw3d(explicit(20*exp(-x^2-y^2)-10,x,0,2,y,-3,3), contour_levels = 15, contour = both, surface_hide = true);
```
Gradient properties

**Theorem:** The gradient $\nabla f(x)$ is perpendicular (orthogonal) to the local tangent (line, plane) of the level curve $L_{=}(f(x))$.

The plot shows the Gradient at different points of the function $f(x_1,x_2)=20 \exp(-x_1^2 - x_2^2)$

Program: wxMaxima:
\[ f1(x1,x2):=20*\exp(-x1^2-x2^2); \]
\[ gx1f1(x1,x2):=\text{diff}(f1(x1,x2),x1,1); \]
\[ gx2f1(x1,x2):=\text{diff}(f1(x1,x2),x2,1); \]
\[ \text{load(dfdraw);} \]
\[ \text{drawdf([gx1f1(x1,x2),gx2f1(x1,x2)],[x1,-2,2],[x2,-2,2]);} \]
Example 1 - Dimension

In one dimension the gradient is given by
\[ \frac{\partial f(x)}{\partial x(x)} = f'(x). \]

\[ \lim_{x \to x_0} f(x) - [f(x_0) + \nabla f(x_0)(x - x_0)] = 0 \]
specializes to
\[ \lim_{x \to x_0} f(x) - [f(x_0) + f'(x_0)(x - x_0)] = 0 \]

This means that locally:
\[ f(x) \approx f(x_0) + f'(x_0)(x - x_0) \]

Now, the left hand side can be written as
\[ \underbrace{f(x_0)}_{=c} - \underbrace{f'(x_0)x_0 + f'(x_0)x}_{=m} \]
and hence \( f(x) \) is locally well approximated by a linear function of the form
\[ f(x) \approx mx + c \]
Single objective, unconstrained

Continuous unconstrained

\[ f(x) \rightarrow \min, \quad x \in \mathbb{R}^n \]

Optimality conditions: \( x \) is a (strict) local optimum, iff

\[ \nabla f(x) = 0, \quad \nabla^2 f(x) \text{ positive (semi-)definite} \]

\[ \nabla f(x) = \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)^\top(x) \]

\[ \nabla^2 f(x) = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j} \right](x)_{i=1,\ldots,n,j=1,\ldots,n} \quad \frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial f}{\partial x_j}(x) \]

A matrix is positive (semi-)definite if all eigenvalues are positive (non-negative)

Often, as in the case \( x^2 + y^2 \rightarrow \min \), a lower bound can be obtained and used to argue whether a (stationary) point is a local/global optimum.
Recall: Simple rules for taking derivatives

\[
\frac{\partial cx}{\partial x} = c \\
\frac{\partial c}{\partial x} = 0 \\
\frac{\partial x^p}{\partial x} = px^{p-1} \\
\frac{\partial \exp(x)}{\partial x} = \exp(x) \\
\frac{\partial u(v(x))}{\partial x} = \frac{\partial u}{\partial x}(v(x))\frac{\partial v}{\partial x}(x) \text{ chain rule} \\
\frac{\partial u(x)v(x)}{\partial x} = \frac{\partial u}{\partial x}(x)v(x) + \frac{\partial v}{\partial x}(x)u(x) \text{ product rule} \\
\frac{\partial \ln(x)}{\partial x} = \frac{1}{x}
\]
Single objective, unconstrained

Necessary condition for optimality:
\[ \nabla f(x) = \left( \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n} \right)^T = (0, \ldots, 0)^T \]

Example:
\[ f(x_1, x_2) = 1.1(x_1)^2 + (x_2)^2 \]
\[ \frac{\partial f}{\partial x_1} = 2.2x_1 = 0 \]
\[ \frac{\partial f}{\partial x_2} = 2x_2 = 0 \]
\[ \Rightarrow x_1 = 0, x_2 = 0. \]

Sufficient condition max.:
\[ \nabla^2 f(x) \text{ positive definite} \]
\[ \nabla^2 f(x) = \begin{pmatrix} 2.2 & 0 \\ 0 & 2 \end{pmatrix}. \]

Eigenvalues are:
\[ \lambda_1 = 2.2 \]
\[ \lambda_2 = 2 \]
\[ \Rightarrow \nabla^2 f(x) \text{ is positive definite in } x \]
\[ \Rightarrow x \text{ is a local maximizer}. \]
Constraints (equalities)

\[ f(x) \rightarrow \min, \text{ s.t. } g_1(x) = 0, \ldots, g_m(x) = 0 \]

All functions are continuously differentiable.

A necessary condition for \( x^* \) to be a local extremum is given, if there exists multipliers \( \lambda_1, \ldots, \lambda_{m+1} \) with at least one \( \lambda_i \neq 0 \) for \( i = 1, \ldots, m + 1 \), such that:

\[
\lambda_1 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_{i+1} \nabla g_i(x^*) = 0
\]

The Lagrange multipliers \( \lambda_i \) are named after Lagrange (1736-1813), who discovered this theorem, but could not prove it. It took 100 years before the proof was found.
Constraints (equalities) - interpretation

Note that $\nabla f(x)$ is perpendicular to the level curves.
Example: One equality constraint, three dimensions

All solutions that satisfy the equality constraints are located on the gray surface.

Level curve of $f$: $f(x_1, x_2, x_3) = \text{const}$
Example: 2 Equality constraints, three dimensions

Points on the intersection of The two planes satisfy both constraints.
Constraints (inequalities)

\[ f(x) \rightarrow \min, \text{ s.t. } g_1(x) \leq 0, \ldots, g_m(x) \leq 0, \text{ all functions are continuously differentiable.} \]

The Karush Kuhn Tucker conditions are said to hold for \( x^* \), if there exists multipliers \( \lambda_1 \geq 0, \ldots, \lambda_{m+1} \geq 0 \) and at least one \( \lambda_i > 0 \) for \( i = 1, \ldots, m + 1 \), such that:

1. \[ \lambda_1 \nabla f(x^*) + \sum_{i=1}^{m} \lambda_{i+1} \nabla g_i(x^*) = 0. \]
2. \[ \lambda_{i+1} g_i(x^*) = 0, i = 1, \ldots, m \]

KKT Theorem - Necessary conditions for smooth, convex programming: Assume the objective and all constraint functions are convex in some \( \epsilon \)-neighborhood of \( x^* \), if \( x^* \) is a local minimum, then there exists \( \lambda_1, \ldots, \lambda_{m+1} \) such that KKT conditions are fulfilled.

Kuhn, US American Mathematician
Constraint (inequality)

As in the case of Lagrange multiplier, we get $m+n$ non-linear equations, the solution of which results in candidate solutions.

The KKT conditions are sufficient for optimiality, provided $\lambda_1 = 1$. In this case $x^*$ is a local minimum.

Note that if $x^*$ is in the interior of the feasible region (a Slater point), all $g_i(x) < 0$ and thus $\lambda_1 > 0$.

[Brinkhuis, Tikhomirov, 2005]
Geometrical interpretation KKT conditions

A constraint function \( g \) is called **active** in \( x \) if \( g(x) = 0 \), i.e. \( x \) is located at the boundary of \( g \).

**Geometrical interpretation of KKT condition:** Let \( a_1, \ldots, a_k \) denote the indexes for the constraint functions that are active in \( x \). Then \( \nabla f(x) \) lies in the cone spanned by \( \nabla g_{a_1}(x), \ldots, \nabla g_{a_k}(x) \).

(Recall: Convex polyhedral cone: \( \{ y \mid y = \sum_{i=1}^{n} \lambda_i d_i \} \) for \( \lambda_i \geq 0 \) and \( d_i \) a set of vectors ’spanning’ the cone.)
Multiobjective Optimization [cf. Miettinnen ‘99]

**Fritz John necessary conditions**

A necessary condition for $x^*$ to be a locally efficient point is that there exists vectors $\lambda_1, \ldots, \lambda_k$ and $\nu_1, \ldots, \nu_m$ such that

(0) $\lambda \succ 0, \nu \succ 0$

(1) $\sum_{i=1}^{k} \lambda_i \nabla f_i(x^*) - \sum_{i=1}^{m} \nu_i \nabla g_i(x^*) = 0.$

(2) $\nu_i g_i(x^*) = 0, i = 1, \ldots, m$

**Karush Kuhn Tucker sufficient conditions for a solution to be Pareto optimal:** Let $x^*$ be a feasible point. Assume that all objective functions are locally convex and all constraint functions are locally concave, and the Fritz John conditions hold in $x^*$, then $x^*$ is a local efficient point.
Unconstrained Multiobjective Optimization

In the unconstrained case Fritz John neccessary conditions reduce to

There exist numbers $\lambda_1, \ldots, \lambda_k$, such that

$$(0) \; \lambda \succ 0$$
$$(1) \; \sum_{i=1}^{k} \lambda_i \nabla f_i(x^*) = 0.$$ 

$x^*$ is optimum for some linear scalarization with some weights $\lambda_1, \ldots, \lambda_k$.

In 2-dimensional spaces this criterion reduces to the observation, that either one of the objectives has a zero gradient (necessary condition for ideal points) or the gradients are parallel.
Strategies to solve multiobjective optimization problems: epsilon-Constraint method

The $\epsilon$-constraint method maximizes $f_1$ while fixing the other objectives $f_2, \ldots, f_m$ to a constant $\epsilon$.

$$f_1(x) \rightarrow \min, f_2(x) = \epsilon_1, \ldots, f_m(x) = \epsilon_{m-1}$$

A grid is used to sample different $\epsilon_i$ values. Selection of grid determines resolution of Pareto front approximation.

In some cases, $\epsilon$ can be eliminated and $f_2$ can be expressed in terms of $f_1$. See example
Strategies to solve multiobjective optimization problems: epsilon-Constraint method

We may find different points by solving single objective problem $U(x) = \sum_{i=1}^{m} w_i f_i(x) \rightarrow \text{min}$, for different positive $w_i$.

This way Pareto optimal solutions will be obtained, as Pareto minima are also minima of $w_1 f_1(x) + w_2 f_2(x)$.

However, we cannot obtain all solutions in concave Pareto fronts this way (see example in Figure below).
Strategies to solve multiobjective optimization problems: Other utility functions

The point where the level curve (or indifference curve) with the best level that has an intersection with the Pareto front is the point that will be obtained when optimizing the utility function.

Depending on the utility function class, some points on PF might not be accessible.
Strategies to solve multiobjective optimization problems: Level set continuation

Recall: KKT conditions, unconstrained case: For efficient $x$ there exists $\lambda_1, \ldots, \lambda_k$, such that $\lambda \succ 0$ and $\sum_{i=1}^{k} \lambda_i \nabla f_i(x^*) = 0$.

Strategy: Solve $m+n$ dimensional equation system with $n+1$ equations:

$$\sum \lambda_i \nabla f_i(x) = 0$$

$$\sum_{i=1}^{m} \lambda_i = 1$$

$$\lambda_i \in \mathbb{R}_0^+, i = 1, \ldots, m$$

Yields $m-1$ dimensional manifold for regular problems. Strategy: Find one point using optimization and extend surface by predictor-corrector method (Hillermeier 2001).
1. Gradient is a vector of first order partial derivatives that is perpendicular to level curves; Hessian contains second order partial derivatives.

2. Local linearization yields optimality condition; in single objective case ‘gradient zero’ and positive/negative definite Hessian.

3. Lagrange multiplier rule can be used to solve constrained optimization problems with equality constraints.

4. KKT conditions generalize it to inequality constraints; negative gradient points in cone spanned by active constraints.

5. KKT conditions for multiobjective optimization require for interior points to be optimal that they have gradients which point in exactly the opposite directions.

6. Different formulations of single objective (un)constrained optimization can be used to find candidates of efficient points.

7. Minimization of linear weighting or utility function yields Pareto optima, but some solutions might not be accessible this way.

8. KKT conditions define equation system the solution of which is an at most m-1 dimensional manifold.
References


