Optimality Conditions for Nonlinear Optimization

GIAN Short Course on Optimization: Applications, Algorithms, and Computation

Sven Leyffer

Argonne National Laboratory

September 12-24, 2016

Outline

1. Preliminaries: Definitions and Notation
2. First-Order Conditions
   - Equality Constrained Nonlinear Programs
   - Inequality Constrained Nonlinear Programs
   - The Karush-Kuhn-Tucker Conditions
3. Second-Order Conditions
   - Second-Order Conditions for Equality Constraints
   - Second-Order Conditions for Inequality Constraints

Preliminaries: Definitions and Notation

Seek optimality conditions for (local) minimizer ...

Definition (Nonlinear Optimization Problem)

$$
\begin{array}{lll}
\text{minimize} & f(x) & \\
\text{subject to} & c_i(x) = 0, & i \in \mathcal{E} \\
& l_i \le c_i(x) \le u_i & i \in \mathcal{I} \\
& l_j \le x_j \le u_j & j = 1, \ldots, n
\end{array}
$$

where

- $f(x)$ and $c_i(x)$ twice continuously differentiable.
- $\mathcal{E}$ indexes equality, $\mathcal{I}$ indexes inequality constraints
- Bounds l_j, u_j, l_i, u_i can be finite or infinite

Also referred to as *nonlinear program (NLP)*.

Often, have additional structure, that can be exploited by solver

Preliminaries: Definitions and Notation

Simplify notation ... other NLPs can be expressed like this.

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & c_i(x) = 0 & i \in \mathcal{E} \\ & c_i(x) \geq 0 & i \in \mathcal{I}. \end{array}$$

Notation

$c_{\mathcal{E}}(x) = 0$, $c_{\mathcal{I}}(x) \geq 0$ denotes equality, inequality constraints.

$$\text{For } \mathcal{E} = \{1, \ldots, m\} \quad c_{\mathcal{E}}(x) = \begin{pmatrix} c_1(x) \\ \vdots \\ c_m(x) \end{pmatrix}$$

Preliminaries: Definitions and Notation

Nonlinear optimization problem

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & c_i(x) = 0 & i \in \mathcal{E} \\ & c_i(x) \geq 0 & i \in \mathcal{I}. \end{array}$$

Definition (Feasible Set)

Feasible set of NLP is set of x that satisfy all constraints

$$\mathcal{F} := \left\{ x | c_{\mathcal{E}}(x) = 0, \text{ and } c_{\mathcal{I}}(x) \geq 0 \right\}$$

Definition (Local and Global Minimizers)

- $x^* \in \mathcal{F}$ is global minimizer, iff $f(x^*) \leq f(x)$ for all $x \in \mathcal{F}$.
- $x^* \in \mathcal{F}$ is local minimizer, iff there exists neighborhood $\mathcal{N}(x^*)$ of x^* such that $f(x^*) \leq f(x)$ for all $x \in \mathcal{F} \cup \mathcal{N}(x^*)$.

Local versus Global Minimizers

Notation.

Gradient of $f(x)$ is $g(x) = \nabla f(x)$, Jacobian of $c(x)$ is $A(x) = \nabla c(x)$.

Remark (Limitations of Optimality Conditions)

- *Optimality conditions only provide local optimality.*
- *Limited to smooth finite-dimensional problems.*
  ... extend to nonsmooth problems using subdifferential $\partial f(x)$

Remark (Importance of Optimality Conditions)

- *Guarantee that candidate solution is local optimum*
- *Indicate when point is not optimal (necessary conditions)*
- *Guide development of optimization methods*

Outline

First-Order Conditions

Theorem (Unconstrained First-Order Conditions)

x^* unconstrained local minimizer $\Rightarrow g^* = 0$.

State this condition equivalently as

$$g^* = 0 \quad \Leftrightarrow \quad s^T g^* = 0, \ \forall s \quad \Leftrightarrow \quad \left\{ s \mid s^T g^* < 0 \right\} = \emptyset,$$

i.e. there are no strict descend directions at x^*

Generalize these conditions

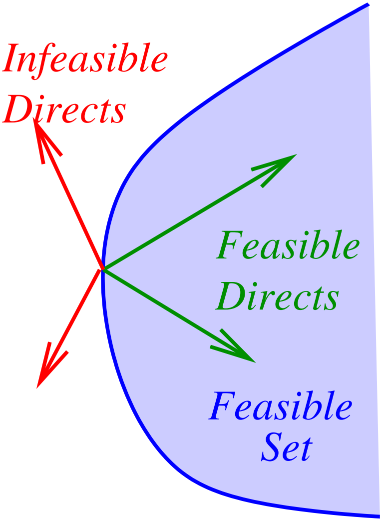
- Must classify feasible directions
- Derive easy-to-check conditions for

$$\left\{ s \mid s^T g^* < 0, \ \forall s \text{ feasible directions} \right\} = \emptyset,$$

i.e. there exist no feasible descend directions.

Concept of Feasible Directions

Feasible directions play central role in optimality ...



Distinguish two cases:

1. Equality constraints only.
2. Inequality constraints.

... equality constraints easier

Equality Constrained Nonlinear Programs

Consider equality constraints only:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & c_{\mathcal{E}}(x) = 0. \end{array}$$

Take infinitesimal step δ from x^*, look at Taylor series expansion:

$$c_i(x^* + \delta) = c_i(x^*) + \delta^T a_i^* + o(\|\delta\|) = \delta^T a_i^* + o(\|\delta\|),$$

because $c_i(x^*) = 0$, where $a_i^* = \nabla c_i(x^*)$

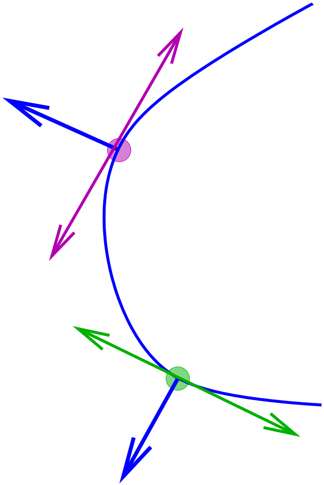
Recall: $a = o(h)$ means $\frac{a}{h} \to 0$ as $h \to 0$

Sufficient Condition for Feasible $x^* + \delta$

$$\delta^T a_i^* + o(\|\delta\|) = 0 \quad \Rightarrow \quad s^T a_i^* = 0 \quad \text{feasible directions}$$

Graphical Interpretation of Feasible Directions

Feasible directions, s such that $s^T a_i^* = 0$ are tangent directions



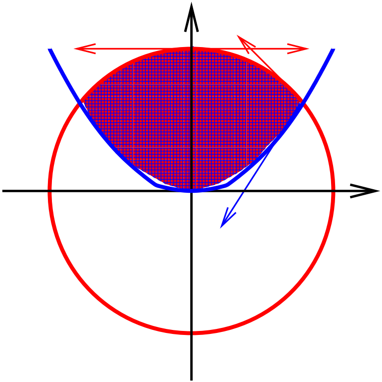
Feasible directions at two different points.

Graphical Interpretation of Feasible Directions

Feasible directions, s such that $s^T a_i^* = 0$ are tangent directions

How to derive feasible directions:

$$\mathcal{F} = \left\{ x \mid x_1^2 - x_2 \le 0,\ x_1^2 + x_2^2 \le 1 \right\}$$



$$\nabla c_1(x) = \begin{pmatrix} 2x_1 \\ -1 \end{pmatrix}$$

$$\nabla c_2(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

At $x = (0;1)$ get $s = (\pm 1; 0)$:

$$(\pm 1; 0)^T (0; 2) = 0$$

At $x = (0.7861; 0.6180)$ get two directions

$$\begin{pmatrix} -0.5367 \\ -0.8438 \end{pmatrix} \quad \begin{pmatrix} -0.6180 \\ 0.7861 \end{pmatrix}$$

Regularity Assumptions

To derive stationarity conditions, need regularity assumption:
"linearized feasible set", looks like nonlinear feasible set

Assumption (Linear Independence of Constraint Normals)

$a_i^* = \nabla c_i(x^*)$, *for* $i = 1, \ldots, m_e$, *are linearly independent.*

An alternative assumption is that all constraints are linear

- Any linearization of a linear constraint is perfect approx.
- Hence, do not need regularity assumptions for LPs and QPs.

Necessary Condition for Equality Constraints

$$\text{minimize } f(x) \quad \text{subject to } c_{\mathcal{E}}(x) = 0$$

Necessary condition: under linear independence assumption:

$$x^* \text{ is a local minimizer } \Rightarrow \left\{ s \mid s^T g^* < 0,\ s^T a_i^* = 0,\ \forall i \in \mathcal{E} \right\} = \emptyset$$

... very difficult to check

Lemma (Necessary Condition for Equality Constraints)

Assume linear independence holds, and x^* is local minimizer, then the following conditions are equivalent:

1. $\left\{ s \mid s^T g^* < 0,\ s^T a_i^* = 0,\ \forall i \in \mathcal{E} \right\} = \emptyset$
2. *There exist Lagrange multipliers, y_i^*, for $i \in \mathcal{E}$ such that*

$$g^* = \sum_{i \in \mathcal{E}} y_i^* a_i^* = A^* y.$$

Graphic Interpretation of FO Conditions

Lemma (Necessary Condition for Equality Constraints)

Assume linear independence holds, and x^* is local minimizer, then the following conditions are equivalent:

1. $\left\{s \mid s^T g^* < 0,\ s^T a_i^* = 0,\ \forall i \in \mathcal{E}\right\} = \emptyset$
2. *There exist Lagrange multipliers, y_i^*, for $i \in \mathcal{E}$ such that*

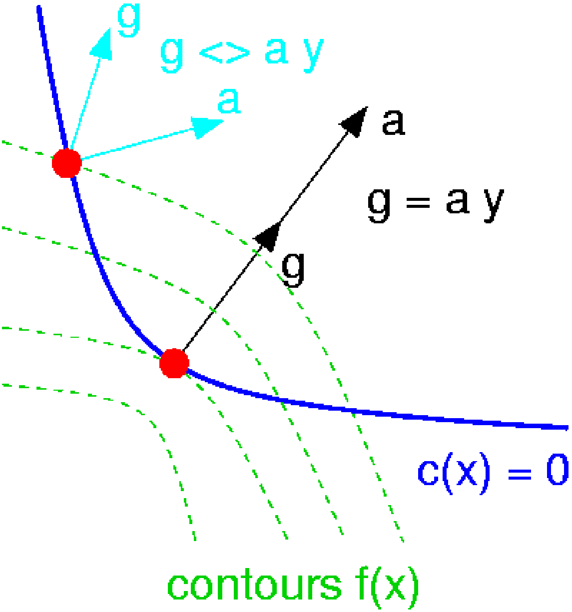
$$g^* = \sum_{i \in \mathcal{E}} y_i^* a_i^* = A^* y.$$

- Can write g^* as linear combination of constraint gradients, a_i^*.
- Linear-independence Assumption implies $\text{rank}(A^*) = m_e$ i.e. A^* has full rank $\Rightarrow$ generalized inverse, A^+, exists

$$y^* = {A^*}^+ g^*, \quad \text{where } {A^*}^+ = \left({A^*}^T A^*\right)^{-1} {A^*}^T,$$

unique multipliers, y^*, also solve $\min \ \|A^* y - g^*\|_2^2$

Graphic Interpretation of FO Conditions



Method of Lagrange Multipliers

Restate conditions in Lemma as system of equations in (x, y):

$$\begin{aligned} g(x) &= A(x)\, y && \text{first-order condition} \\ c(x) &= 0 && \text{feasibility.} \end{aligned}$$

Define Lagrangian function, $\mathcal{L}(x, y) := f(x) - y^T c(x)$

Method of Lagrange Multipliers

First-order optimality conditions equivalent to

$$\nabla_x \mathcal{L}(x, y) = 0, \quad \text{and} \quad \nabla_y \mathcal{L}(x, y) = 0.$$

Can apply Newton's method to nonlinear system in (x, y)

Finding stationary points $\Leftrightarrow$ finding stationary point of Lagrangian

Effect of Perturbations: Sensitivity Analysis

Express effect of perturbation to constraint, $c_i(x) = \epsilon_i$ on optimum
Let $x(\epsilon)$ and $y(\epsilon)$ denote optimal values after perturbation

$$f(x(\epsilon)) = \mathcal{L}(x(\epsilon), y(\epsilon)) = f(x(\epsilon)) - y(\epsilon)^T (c(x) - \epsilon)$$

Chain rule implies

$$\frac{df}{d\epsilon_i} = \frac{d\mathcal{L}}{d\epsilon_i} = \frac{\partial x^T}{\partial \epsilon_i} \nabla_x \mathcal{L} + \frac{\partial y^T}{\partial \epsilon_i} \nabla_y \mathcal{L} + \frac{\mathcal{L}}{\partial \epsilon_i}$$

Observe, that $\nabla_x \mathcal{L}(x, y) = 0$ and $\nabla_y \mathcal{L}(x, y) = 0$, hence

$$\frac{\mathcal{L}}{\partial \epsilon_i} = -y_i \;\Rightarrow\; \frac{df}{d\epsilon_i} = -y_i .$$

Sensitivity Interpretation of Multipliers

Multiplier, y_i, gives rate of change in objective to perturbation right-hand-side of constraint i.

Inequality Constrained Nonlinear Programs

Now consider both equality and inequality constraints

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & c_i(x) = 0 & i \in \mathcal{E} \\ & c_i(x) \geq 0 & i \in \mathcal{I}. \end{array}$$

only need to consider active constraints

$$\mathcal{A}^* := \mathcal{A}(x^*) := \{i \in \mathcal{E} \cup \mathcal{I} \mid c_i(x^*) = 0\} \qquad \text{active set}.$$

... includes all equality constraints

Again, looking for feasible directions ... now for inequalities.

Inequality Constrained Nonlinear Programs

Now consider both equality and inequality constraints

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & c_i(x) = 0 & i \in \mathcal{E} \\ & c_i(x) \geq 0 & i \in \mathcal{I}. \end{array}$$

Let δ be small incremental step for active inequality, $i \in \mathcal{I} \cap \mathcal{A}^*$:

$$c_i(x^* + \delta) = c_i(x^*) + \delta^T a_i^* + o(\|\delta\|) = \delta^T a_i^* + o(\|\delta\|).$$

Now require step to remain feasible only wrt one side:

$$c_i(x^* + \delta) \geq 0 \;\Leftrightarrow\; \delta^T a_i^* + o(\|\delta\|)$$

Hence, δ lies in direction s:

feasible directions $\quad s^T a_i^* \geq 0, \; \forall i \in \mathcal{I} \cap \mathcal{A}^*, \quad s^T a_i^* = 0, \; \forall i \in \mathcal{E}.$

... again need a regularity assumption ...

Regularity Assumption for Inequality Constraints

Need regularity assumption to ensure that linearized analysis captures nonlinear geometry

Assumption (Linear Independence Constraint Qualification)

The linear-independence constraint qualification (LICQ) holds at x^* for the NLP, iff $a_i^* = \nabla c_i(x^*)$, for $i \in \mathcal{A}^*$, are linearly independent.

The next assumption is slightly weaker, and implies the LICQ.

Assumption (Mangasarian-Fromowitz Constraint Qualification)

The Mangasarian-Fromowitz constraint qualification (MFCQ) holds at x^* for the NLP, iff $a_i^* = \nabla c_i(x^*)$, for $i \in \mathcal{E}$, are linearly independent, and there exists $s \neq 0$ such that

$$s^T a_i^* > 0, \; \forall i \in \mathcal{I} \cap \mathcal{A}^*.$$

Why We Need Regularity Assumptions

Consider the NLP

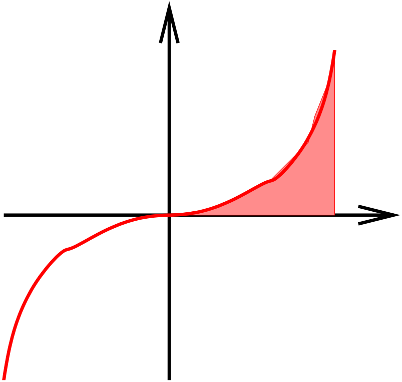
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & x_1 \\ \text{subject to} \quad & x_2 \le x_1^3 \\ & x_2 \ge 0 \end{aligned}$$

Has optimum at cusp

$$x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

... but constraints violate MFCQ
$\Rightarrow$ bogus "feasible" direction

$$s = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$$



MFCQ fails at cusp, $x = 0$

Regularity Assumption for Inequality Constraints

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & c_i(x) = 0 & i \in \mathcal{E} \\ & c_i(x) \geq 0 & i \in \mathcal{I}. \end{array}$$

Assumption (Mangasarian-Fromowitz Constraint Qualification)

The Mangasarian-Fromowitz constraint qualification (MFCQ) holds at x^* for the NLP, iff $a_i^* = \nabla c_i(x^*)$, for $i \in \mathcal{E}$, are linearly independent, and there exists $s \neq 0$ such that

$$s^T a_i^* > 0, \; \forall i \in \mathcal{I} \cap \mathcal{A}^*.$$

MFCQ is stronger than needed:

$$\left\{ s | s^T g^* < 0, \; s^T a_i^* = 0, \; \forall i \in \mathcal{E}, \; s^T a_i^* \geq 0, \; \forall i \in \mathcal{I} \cap \mathcal{A}^* \right\} = \emptyset$$

... but this condition really difficult to check.

Necessary Condition for Nonlinear Optimization

Lemma (First-Order Conditions for Optimality)

Assume that LICQ or MFCQ hold, and that x^* is local minimizer, then the following two conditions are equivalent:

1. *There exist no feasible descend direction:*

$$\left\{ s | s^T g^* < 0,\ s^T a_i^* = 0,\ \forall i \in \mathcal{E},\ s^T a_i^* \geq 0,\ \forall i \in \mathcal{I} \cap \mathcal{A}^* \right\} = \emptyset$$

2. *There exist so-called Lagrange multipliers, y_i^*, for $i \in \mathcal{A}^*$:*

$$g^* = \sum_{i \in \mathcal{A}^*} y_i^* a_i^* = A^* y \quad \text{where } y_i^* \geq 0,\ \forall i \in \mathcal{I} \cap \mathcal{A}^*.$$

Necessary Condition for Nonlinear Optimization

$$g^* = \sum_{i \in \mathcal{A}^*} y_i^* a_i^* = A^* y \quad \text{where } y_i^* \geq 0, \ \forall i \in \mathcal{I} \cap \mathcal{A}^*.$$

Remark (Towards an Algorithms for NLP)

Assume at non-stationary point with

- *Multiplier* $y_q < 0$ *for some* $q \in \mathcal{I}$
  e.g. least-squares multiplier
- *Have direction* s *with* $s^T a_q = 1$

Then reduce objective by step in this feasible direction s*.*

Basis for active-set methods for linear and quadratic programming!

The Karush-Kuhn-Tucker Conditions

$$\begin{array}{lll} \text{minimize} & f(x) & \\ \text{subject to} & c_i(x) = 0 & i \in \mathcal{E} \\ & c_i(x) \geq 0 & i \in \mathcal{I}. \end{array}$$

Theorem (Karush-Kuhn-Tucker (KKT) Conditions)

x^* local minimizer of NLP and assume LICQ or MFCQ hold at x^*. Then there exist Lagrange multipliers, y^* such that

$$\nabla_x \mathcal{L}(x^*, y^*) = 0 \qquad \textit{first order condition} \qquad (1)$$

$$c_{\mathcal{E}}(x^*) = 0 \qquad \textit{feasibility} \qquad (2)$$

$$c_{\mathcal{I}}(x^*) \geq 0 \qquad \textit{feasibility} \qquad (3)$$

$$y^*_{\mathcal{I}} \geq 0 \qquad \textit{dual feasibility} \qquad (4)$$

$$y^*_i c_i(x^*) = 0 \qquad \textit{complementary slackness.} \qquad (5)$$

Interpretation of KKT Conditions

Remark (Stationarity Conditions and Algorithms)

Take standard NLP & linearize about stationary point, x^*, then: KKT conditions are the FO conditions of linearized problem:

$$\begin{aligned} \underset{d}{\mathit{minimize}} \quad & f(x^*) + d^T \nabla f(x^*) \\ \mathit{subject\ to}\ & c_i(x^*) + d^T \nabla c_i(x^*) = 0, \;\; i \in \mathcal{E} \\ & c_i(x^*) + d^T \nabla c_i(x^*) \geq 0, \;\; i \in \mathcal{I}, \end{aligned}$$

- Motivates algorithms such as SLP, SQP, SLQP, SQQP, ...
- Extends FO conditions to structured NLP, e.g. MPECs, ...
  ... and hence defines new structured algorithmic approaches

Outline

Second-Order Conditions

KKT conditions are first-order necessary conditions.

Goal

Extend second-order from the unconstrained case

Remark

Important to include second-order effects from constraints

- *Can replace objective:*

$$\underset{x}{\text{minimize}}\ f(x) \quad \Leftrightarrow \quad \underset{x,\eta}{\text{minimize}}\ \eta \quad \text{subject to}\ \eta \geq f(x)$$

- *Need to consider* $\nabla^2 c_i(x)$*, not just* $\nabla^2 f(x)$.

Again convenient to distinguish equality and inequality constraints.

Second-Order Conditions for Equality Constraints

Let x^* is KKT point, and a_i^* for $i \in \mathcal{E}$ linearly independent
Let δ be an incremental step along feasible direction, s.

$$
\begin{aligned}
f(x^* + \delta) &= \mathcal{L}(x^* + \delta, y^*) \\
&= \mathcal{L}(x^*, y^*) + \delta^T \nabla_x \mathcal{L}(x^*, y^*) + \frac{1}{2}\delta^T W^* \delta + o(\|\delta\|^2) \\
&= f(x^*) + \frac{1}{2}\delta^T W^* \delta + o(\|\delta\|^2),
\end{aligned}
$$

where Hessian of Lagrangian is:

$$
W^* = \nabla^2 \mathcal{L}(x^*, y^*) = \nabla^2 f(x^*) + \sum_{i \in \mathcal{E}} y_i^* \nabla^2 c_i(x^*)
$$

Optimality of x^* implies

$$
s^T W^* s \geq 0, \ \forall s : s^T a_i^* = 0.
$$

i.e. Lagrangian has nonnegative curvature for all feasible directions

Second-Order Conditions for Equality Constraints

Proposition (Second-Order Necessary Condition)

x^* local minimizer, and if constraint qualification holds, then

$$s^T \nabla^2 \mathcal{L}(x^*, y^*) s \geq 0 \; \forall s : s^T a_i^* = 0.$$

Can also state sufficient condition for local minimizer.

Proposition (Second-Order Sufficient Condition)

If $\nabla_x \mathcal{L}(x^*, y^*) = 0$, if $c(x^*) = 0$, and if

$$s^T \nabla^2 \mathcal{L}(x^*, y^*) s > 0, \; \forall s \neq 0 : s^T a_i^* = 0,$$

then x^* is a local minimizer.

Note: $\exists$ gap between necessary and sufficient conditions.

Second-Order Conditions for Inequality Constraints

To derive second-order conditions consider active constraints, $\mathcal{A}^*$.
$\Rightarrow$ NLP equivalent to equality NLP, if $y_i^* > 0, \ \forall i \in \mathcal{I} \cap \mathcal{A}^*$,

Simplifying Assumption

Assume strict complementarity: $y_i^* > 0, \ \forall i \in \mathcal{I} \cap \mathcal{A}^*$,

Proposition (Second-Order Sufficient Condition)

If $\nabla_x \mathcal{L}(x^*, y^*) = 0$, *if* $c(x^*) = 0$, *if strict complementarity holds, i.e.* $y_i^* > 0, \ \forall i \in \mathcal{I} \cap \mathcal{A}^*$, *and if*

$$s^T \nabla^2 \mathcal{L}(x^*, y^*) s > 0, \ \forall s \neq 0 : s^T a_i^* = 0, \ \forall i \in \mathcal{A}^*,$$

then it follows that x^* *is a local minimizer.*

Second-Order Conditions for Inequality Constraints

More rigorous results without strict complementarity possible ...
... needs Hessian $\nabla^2 \mathcal{L}$ positive definite over cone impractical

Check sufficient conditions by finding inertia of KKT matrix,

$$\begin{bmatrix} W^* & A^* \\ {A^*}^T & 0 \end{bmatrix}.$$

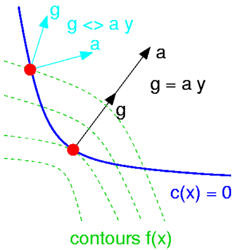
Theorem

If inertia of KKT matrix is $[n-m, 0, m]$, then second order conditions are satisfied, where $m = |\mathcal{A}^*|$.

KKT matrix with inertia is $[n-m, 0, m]$ is second-order sufficient

Matrix inertia: triple of positive, zero, and negative eigenvalues.

Summary and Take-Aways



Derived Optimality Conditions for NLPs

- Intuitive geometric interpretation
- Motivate algorithmic approaches (soon)

Optimality Conditions Require Regularity

- Not easy to check a priori (LICQ is OK)
- What happens if regularity does not hold?
- Algorithms often detect lack of regularity
  ... fail "gracefully" ...