KKT conditions and Duality

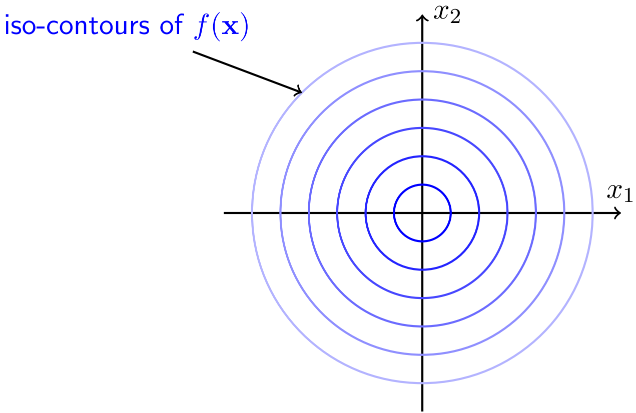
March 23, 2012

Want to solve this constrained optimization problem

$$\boxed{\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) = \min_{\mathbf{x}\in\mathbb{R}^2} .4\,(x_1^2 + x_2^2)}$$

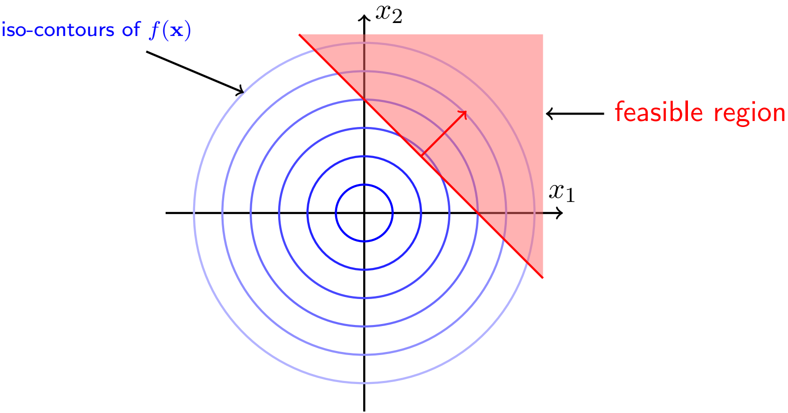
subject to

$$\boxed{g(\mathbf{x}) = 2 - x_1 - x_2 \leq 0}$$



$$\boxed{f(\mathbf{x}) = .4\,(x_1^2 + x_2^2)}$$

Tutorial example - Constraint



$$\boxed{g(\mathbf{x}) = 2 - x_1 - x_2 \leq 0}$$

Solve this problem with Lagrange Multipliers

Can solve this constrained optimization with Lagrange multipliers:

$$\boxed{L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda\, g(\mathbf{x})}$$

Solution:

The Lagrangian is

$$\mathcal{L}(\mathbf{x}, \lambda) = .4\, x_1^2 + .4\, x_2^2 + \lambda\, (2 - x_1 - x_2)$$

The KKT conditions say that at an optimum $\lambda^* \geq 0$ and

$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \lambda^*)}{\partial x_1} = .8\, x_1^* - \lambda^* = 0$$

$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \lambda^*)}{\partial x_2} = .8\, x_2^* - \lambda^* = 0$$

$$\frac{\partial \mathcal{L}(\mathbf{x}^*, \lambda^*)}{\partial \lambda} = 2 - x_1^* - x_2^* = 0$$

Solve this problem with Lagrange Multipliers

Can solve this constrained optimization with Lagrange multipliers:

$$\boxed{L(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda\, g(\mathbf{x})}$$

Solution ctd:

Find $(x_1^*, x_2^*, \lambda^*)$ which fulfill these simultaneous equations. The first two equations imply

$$x_1^* = \frac{5}{4}\lambda^*, \qquad x_2 = \frac{5}{4}\lambda^*$$

Substituting these into the last equation we get

$$8 - 5\lambda^* - 5\lambda^* = 0 \quad \Longrightarrow \lambda^* = \frac{4}{5} \leftarrow \text{greater than 0}$$

and in turn this means

$$x_1^* = \frac{5}{4}\lambda^* = 1, \qquad x_2^* = \frac{5}{4}\lambda^* = 1$$

Solve this particular problem in another way

Alternate solution:

Construct the *Lagrangian dual function*

$$q(\lambda) = \min_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda) = \min_{\mathbf{x}} \; (f(\mathbf{x}) + \lambda g(\mathbf{x}))$$

Find optimal value of $\mathbf{x}$ wrt $\mathcal{L}(\mathbf{x}, \lambda)$ in terms of the Lagrange multiplier:

$$x_1^* = \frac{5}{4}\lambda, \qquad\qquad x_2^* = \frac{5}{4}\lambda$$

Substitute back into the expression of $\mathcal{L}(\mathbf{x}, \lambda)$ to get

$$q(\lambda) = \frac{5}{4}\lambda^2 + \lambda\,(2 - \frac{5}{4}\lambda - \frac{5}{4}\lambda)$$

Find $\lambda \geq 0$ which maximizes $q(\lambda)$. Luckily in this case the global optimum of $q(\lambda)$ corresponds to the constrained optimum

$$\frac{\partial q(\lambda)}{\partial \lambda} = -\frac{5}{2}\lambda + 2 = 0 \quad \Longrightarrow \quad \lambda^* = \frac{4}{5} \quad \Longrightarrow \quad x_1^* = x_2^* = 1$$

Solve the same problem in another way

The Primal Problem

$$\min_{\mathbf{x}\in\mathbb{R}^2} f(\mathbf{x}) \quad \text{subject to} \quad g(\mathbf{x}) \leq 0$$

The **Lagrangian Dual Problem**

$$\boxed{\max_{\lambda\in\mathbb{R}} q(\lambda) \quad \text{subject to} \quad \lambda \geq 0}$$

where

$$\boxed{q(\lambda) = \min_{\mathbf{x}\in\mathbb{R}^2} \left(f(\mathbf{x}) + \lambda\, g(\mathbf{x}) \right)}$$

is referred to as the *Lagrangian dual function*.

In general we will have multiple inequality and equality constraints. The statement of the **Primal Problem** is

$$\boxed{\min_{\mathbf{x}\in X} f(\mathbf{x})}$$

subject to

$$\boxed{\mathbf{g}(\mathbf{x}) \leq \mathbf{0} \quad \text{and} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}}$$

Lagrangian Dual Problem

$$\boxed{\max_{\boldsymbol{\lambda},\boldsymbol{\mu}} q(\boldsymbol{\lambda},\boldsymbol{\mu}) \;\; \text{subject to} \;\; \boldsymbol{\lambda} \geq \mathbf{0}}$$

where

$$\boxed{q(\boldsymbol{\lambda},\boldsymbol{\mu}) = \min_{\mathbf{x}} \left[f(\mathbf{x}) + \boldsymbol{\lambda}^t \, \mathbf{g}(\mathbf{x}) + \boldsymbol{\mu}^t \, \mathbf{h}(\mathbf{x}) \right]}$$

is the *Lagrangian dual function*.

This dual approach is **not guaranteed to succeed**. However,

- It does for a certain class of functions
- In these cases it often leads to a simpler optimization problem.
- Particularly in the case when the dimension of $\mathbf{x}$ is much larger than the number of constraints.
- The expression of $\mathbf{x}^*$ in terms of the Lagrange multipliers may give some insight into the optimal solution i.e. the optimal separating hyper-plane found by the SVM.

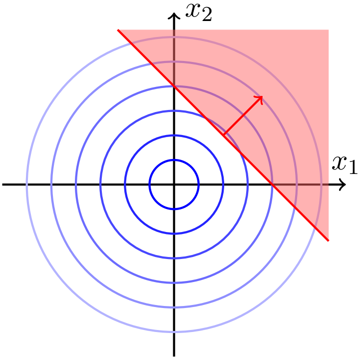
This dual approach is **not guaranteed to succeed**. However,

- It does for a certain class of functions
- In these cases it often leads to a simpler optimization problem.
- Particularly in the case when the dimension of $\mathbf{x}$ is much larger than the number of constraints.
- The expression of $\mathbf{x}^*$ in terms of the Lagrange multipliers may give some insight into the optimal solution i.e. the optimal separating hyper-plane found by the SVM.

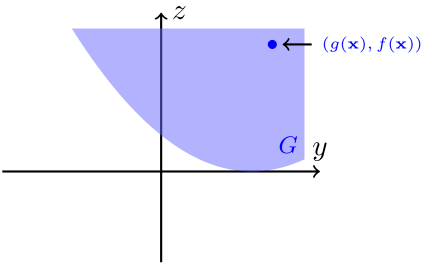
We will now focus on the geometry of the dual solution...

Geometry of the Dual Problem

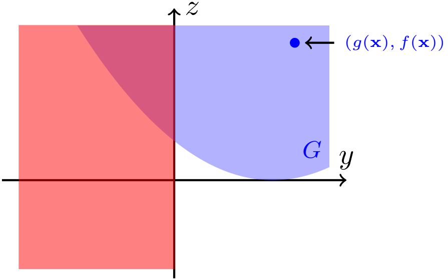
Map the original problem



$\Rightarrow$



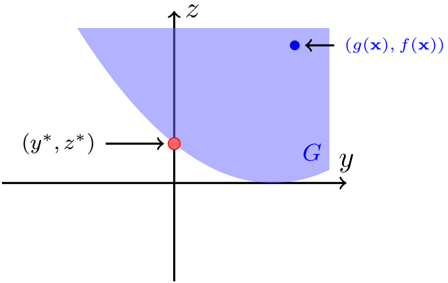
- Map each point $\mathbf{x} \in \mathbb{R}^2$ to $(g(\mathbf{x}), f(\mathbf{x})) \in \mathbb{R}^2$.
- This map defines the set
$$G = \{(y, z) \mid y = g(\mathbf{x}),\ z = f(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{R}^2\}.$$
- **Note:** $\mathcal{L}(\mathbf{x}, \lambda) = z + \lambda\, y$ for some z and y.



Define $G \subset \mathbb{R}^2$ as the image of $\mathbb{R}^2$ under the (g, f) map

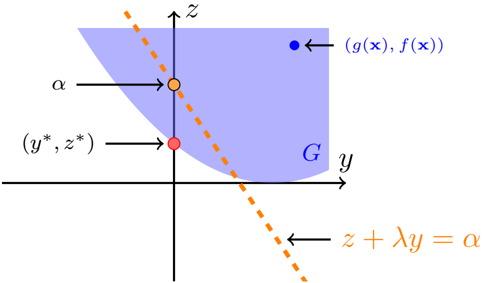
$$\boxed{G = \{(y, z) \mid y = g(\mathbf{x}),\ z = f(\mathbf{x}) \text{ for some } \mathbf{x} \in \mathbb{R}^2\}}$$

In this space only points with $y \leq 0$ correspond to feasible points.



- The primal problem consists in finding a point in G with $y \leq 0$ that has minimum ordinate z.
- Obviously this optimal point is (y^*, z^*).

Visualization of the Lagrangian



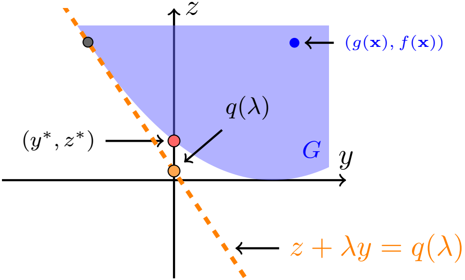
- Given a $\lambda \geq 0$, the *Lagrangian* is given by

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda g(\mathbf{x}) = z + \lambda\, y$$

  with $(y, z) \in G$.

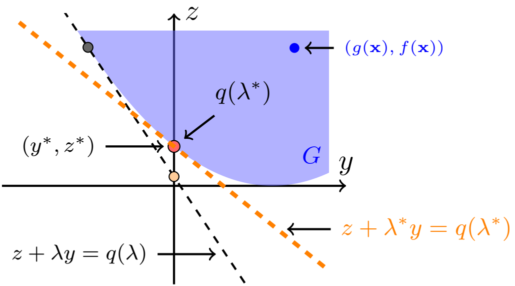
- Note $z + \lambda y = \alpha$ is the eqn of a straight line with slope $-\lambda$ that intercepts the z-axis at α.

Visualization of the Lagrangian Dual function



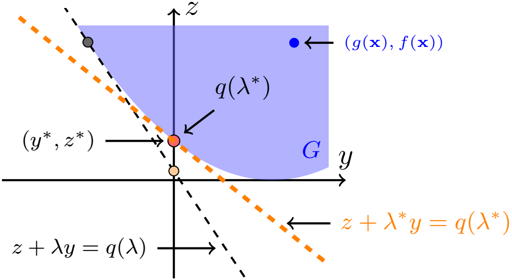
For a given $\lambda \geq 0$ *Lagrangian dual sub-problem* is find: $\min_{(y,z)\in G} (z + \lambda\, y)$

- Move the line $z + \lambda y$ in the direction $(-\lambda, -1)$ while remaining in contact with G.
- The last intercept on the z-axis obtained this way is the value of $q(\lambda)$ corresponding to the given $\lambda \geq 0$.



Finally want to find the dual optimum: $\max_{\lambda} q(\lambda)$

- the line with slope $-\lambda$ with maximal intercept, $q(\lambda)$, on the z-axis.
- This line has slope λ^* and dual optimal solution $q(\lambda^*)$.



- For this problem the optimal dual objective z^* equals the optimal primal objective z^*.
- In such cases, there is **no duality gap (strong duality)**.

Properties of the Lagrangian Dual Function

Theorem

Let $D_q = \{\boldsymbol{\lambda} \,|\, q(\boldsymbol{\lambda}) > -\infty\}$ *then* $q(\boldsymbol{\lambda})$ *is concave function on* D_q*.*

Proof.

For any $\mathbf{x} \in X$ and $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in D_q$ and $\alpha \in (0,1)$

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2) &= f(\mathbf{x}) + (\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2)^t g(\mathbf{x}) \\ &= \alpha(f(\mathbf{x}) + \boldsymbol{\lambda}_1^t g(\mathbf{x})) + (1-\alpha)(f(\mathbf{x}) + \boldsymbol{\lambda}_2^t g(\mathbf{x})) \\ &= \alpha\,\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1) + (1-\alpha)\,\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_2).\end{aligned}$$

Take the $\min$ on both sides

$$\begin{aligned}\min_{\mathbf{x}\in X}\{\mathcal{L}(\mathbf{x}, \alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2)\} &= \min_{\mathbf{x}\in X}\{\alpha\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1) + (1-\alpha)\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_2)\} \\ &\geq \alpha \min_{\mathbf{x}\in X}\{\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_1)\} + (1-\alpha) \min_{\mathbf{x}\in X}\{\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}_2)\}\end{aligned}$$

Therefore

$$q(\alpha\boldsymbol{\lambda}_1 + (1-\alpha)\boldsymbol{\lambda}_2) \geq \alpha\, q(\boldsymbol{\lambda}_1) + (1-\alpha)\, q(\boldsymbol{\lambda}_2)$$

This implies that q is concave over D_q. $\square$

The set of Lagrange Multipliers is convex

Theorem

Let $D_q = \{\boldsymbol{\lambda} \,|\, q(\boldsymbol{\lambda}) > -\infty\}$. This constraint ensures valid Lagrange Multipliers exist. Then D_q is a convex set.

Proof.

Let $\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \in D_q$. Therefore $q(\boldsymbol{\lambda}_1) > -\infty$ and $q(\boldsymbol{\lambda}_2) > -\infty$. Let $\alpha \in (0, 1)$, then as q is concave

$$q(\alpha\,\boldsymbol{\lambda}_1 + (1-\alpha)\,\boldsymbol{\lambda}_2) \geq \alpha\, q(\boldsymbol{\lambda}_1) + (1-\alpha)\, q(\boldsymbol{\lambda}_2) > -\infty$$

and this implies

$$\alpha\,\boldsymbol{\lambda}_1 + (1-\alpha)\,\boldsymbol{\lambda}_2 \in D_q$$

Hence D_q is a convex set. $\square$

Significance of these results

- The dual is always concave, irrespective of the primal problem.
- Therefore finding the **optimum of the dual function** is a **convex optimization problem**.

Weak Duality

Theorem (Weak Duality)

Let $\mathbf{x}$ *be a feasible solution,* $\mathbf{x} \in \mathcal{X}, g(\mathbf{x}) \leq 0$ *and* $h(\mathbf{x}) = 0$*, to the primal problem* P*. Let* $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ *be a feasible solution,* $\boldsymbol{\lambda} \geq 0$*, to the dual problem* D*. Then*

$$f(\mathbf{x}) \geq q(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

Proof of the Weak Duality Theorem.

Remember

$$q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf\{f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{i=1}^{l} \mu_i h_i(\mathbf{x}) \,:\, \mathbf{x} \in X_F\}$$

Then we have

$$\begin{aligned} q(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf\{f(\tilde{\mathbf{x}}) + \boldsymbol{\lambda}^t g(\tilde{\mathbf{x}}) + \boldsymbol{\mu}^t h(\tilde{\mathbf{x}}) \,:\, \tilde{\mathbf{x}} \in X_F\} \\ &\leq f(\mathbf{x}) + \boldsymbol{\lambda}^t g(\mathbf{x}) + \boldsymbol{\mu}^t h(\mathbf{x}) \\ &\leq f(\mathbf{x}) \end{aligned}$$

and the result follows. $\square$

Corollary

Let

$$f^* = \inf\{f(\mathbf{x}) \,:\, \mathbf{x} \in X, g(\mathbf{x}) \geq 0, h(\mathbf{x}) = 0\}$$
$$q^* = \sup\{q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \,:\, \boldsymbol{\lambda} \geq 0\}$$

then

$$\boxed{q^* \leq f^*}$$

- Thus the

  optimal value of the primal problem $\geq$ optimal value of the dual problem.
- If optimal value of the primal problem $>$ optimal value of the dual problem, then there exists a **duality gap**.

Corollary

Let

$$f^* = \inf\{f(\mathbf{x}) \,:\, \mathbf{x} \in X, g(\mathbf{x}) \geq 0, h(\mathbf{x}) = 0\}$$
$$q^* = \sup\{q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \,:\, \boldsymbol{\lambda} \geq 0\}$$

then

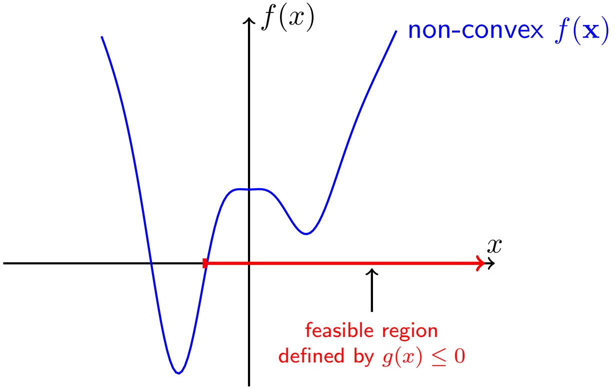
$$\boxed{q^* \leq f^*}$$

- Thus the

  optimal value of the primal problem $\geq$ optimal value of the dual problem.
- If optimal value of the primal problem $>$ optimal value of the dual problem, then there exists a **duality gap**.

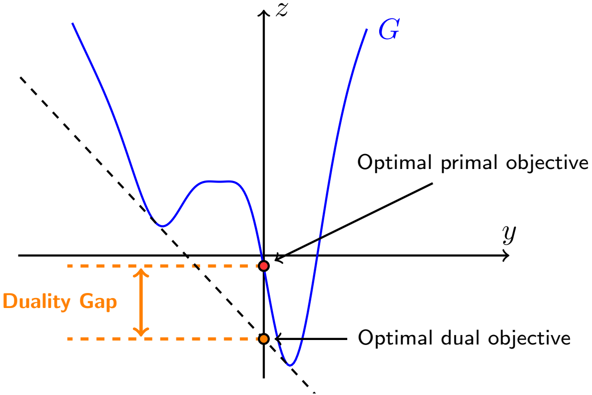
Example with a Duality Gap

Example with a non-convex objective function



- Consider the constrained optimization of this 1D non-convex objective function.
- Let's visualize $G = \{(y, z) \mid \exists x \in \mathbb{R} \text{ s.t. } y = g(x),\, z = f(x))\}$ and its dual solution...

Dual Solution $\leq$ Primal Solution: Have a Duality Gap



- Above is the geometric interpretation of the primal and dual problems.
- Note there exists a **duality gap** due to the nonconvexity of the set G.

Strong Duality

When does Dual Solution = Primal Solution?

The **Strong Duality Theorem** states, that if some suitable convexity conditions are satisfied, then there is no duality gap between the primal and dual optimisation problems.

Theorem (Strong Duality)

Let

- X *be a non-empty convex set in* $\mathbb{R}^n$
- $f : X \to \mathbb{R}$ *and each* $g_i : \mathbb{R}^n \to \mathbb{R}$ *(*$i = 1, \ldots, m$*) be convex,*
- *each* $h_i : \mathbb{R}^n \to \mathbb{R}$ *(*$i = 1, \ldots, l$*) be affine.*

If

- *there exists* $\hat{\mathbf{x}} \in X$ *such that* $g(\hat{\mathbf{x}}) < 0$ *and*
- $\mathbf{0} \in \operatorname{int}(\mathbf{h}(X))$ *where* $\mathbf{h}(X) = \{\mathbf{h}(\mathbf{x}) \,:\, \mathbf{x} \in X\}$.

then

$$\inf\{f(\mathbf{x}) \,:\, \mathbf{x} \in X, g(\mathbf{x}) \leq 0, h(\mathbf{x}) = 0\} = \sup\{q(\boldsymbol{\lambda}, \boldsymbol{\mu}) \,:\, \boldsymbol{\lambda} \geq \mathbf{0}\}$$

where $q(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf\{f(\mathbf{x}) + \boldsymbol{\lambda}^t \mathbf{g}(\mathbf{x}) + \boldsymbol{\mu}^t \mathbf{h}(\mathbf{x}) \,:\, \mathbf{x} \in X\}$.

Theorem (Strong Duality ctd)

Furthermore, if

$$\inf\{f(\mathbf{x}) \,:\, \mathbf{x} \in X, g(\mathbf{x}) \leq 0, h(\mathbf{x}) = 0\} > -\infty$$

then the

$$\sup\{q(\boldsymbol{\lambda}, \boldsymbol{\mu}) : \boldsymbol{\lambda} \geq 0\}$$

is achieved at $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ *with* $\boldsymbol{\lambda}^* \geq 0$*. If the* $\inf$ *is achieved at* $\mathbf{x}^*$ *then*

$$(\boldsymbol{\lambda}^*)^t \mathbf{g}(\mathbf{x}^*) = 0$$