Constrained Optimization Using Lagrange Multipliers

CEE 201L. Uncertainty, Design, and Optimization

Department of Civil and Environmental Engineering
Duke University
Henri P. Gavin and Jeffrey T. Scruggs
Spring 2020

In optimal design problems, values for a set of n design variables, $(x_1, x_2, \cdots x_n)$, are to be found that minimize a scalar-valued objective function of the design variables, such that a set of m inequality constraints, are satisfied. Constrained optimization problems are generally expressed as

$$\min_{x_1,x_2,\cdots,x_n} \; J = f(x_1, x_2, \cdots, x_n) \qquad \text{such that} \qquad \begin{array}{c} g_1(x_1, x_2, \cdots, x_n) \le 0 \\ g_2(x_1, x_2, \cdots, x_n) \le 0 \\ \vdots \\ g_m(x_1, x_2, \cdots, x_n) \le 0 \end{array} \tag{1}$$

If the objective function is quadratic in the design variables and the constraint equations are linearly independent, the optimization problem has a unique solution.

Consider the simplest constrained minimization problem:

$$\min_x \; \frac{1}{2}kx^2 \qquad \text{where} \qquad k > 0 \qquad \text{such that} \qquad x \ge b \; . \tag{2}$$

This problem has a single design variable, the objective function is quadratic ($J = \frac{1}{2}kx^2$), there is a single constraint inequality, and it is linear in x ($g(x) = b - x$). If $g > 0$, the constraint equation constrains the optimum and the optimal solution, x^*, is given by $x^* = b$. If $g \le 0$, the constraint equation does not constrain the optimum and the optimal solution is given by $x^* = 0$. Not all optimization problems are so easy; most optimization methods require more advanced methods. The methods of Lagrange multipliers is one such method, and will be applied to this simple problem.

Lagrange multiplier methods involve the modification of the objective function through the addition of terms that describe the constraints. The objective function $J = f(x)$ is *augmented* by the constraint equations through a set of non-negative multiplicative *Lagrange multipliers*, $\lambda_j \ge 0$. The augmented objective function, $J_A(x)$, is a function of the n design variables and m Lagrange multipliers,

$$J_A(x_1, x_2, \cdots, x_n, \; \lambda_1, \lambda_2, \cdots, \lambda_m) = f(x_1, x_2, \cdots, x_n) + \sum_{j=1}^{m} \lambda_j \; g_j(x_1, x_2, \cdots, x_n) \tag{3}$$

For the problem of equation (2), $n = 1$ and $m = 1$, so

$$J_A(x, \lambda) = \frac{1}{2}kx^2 + \lambda \; (b - x) = \frac{1}{2}kx^2 - \lambda x + \lambda b \tag{4}$$

The Lagrange multiplier, λ, serves the purpose of modifying (augmenting) the objective function from one quadratic ($\frac{1}{2}kx^2$) to another quadratic ($\frac{1}{2}kx^2 - \lambda x + \lambda b$) so that the minimum of the modified quadratic satisfies the constraint ($x \ge b$).

Case 1: $b = 1$

If $b = 1$ then the minimum of $\frac{1}{2}kx^2$ is constrained by the inequality $x \geq b$, and the optimal value of λ should minimize $J_A(x, \lambda)$ at $x = b$. Figure 1(a) plots $J_A(x, \lambda)$ for a few non-negative values of λ and Figure 1(b) plots contours of $J_A(x, \lambda)$.

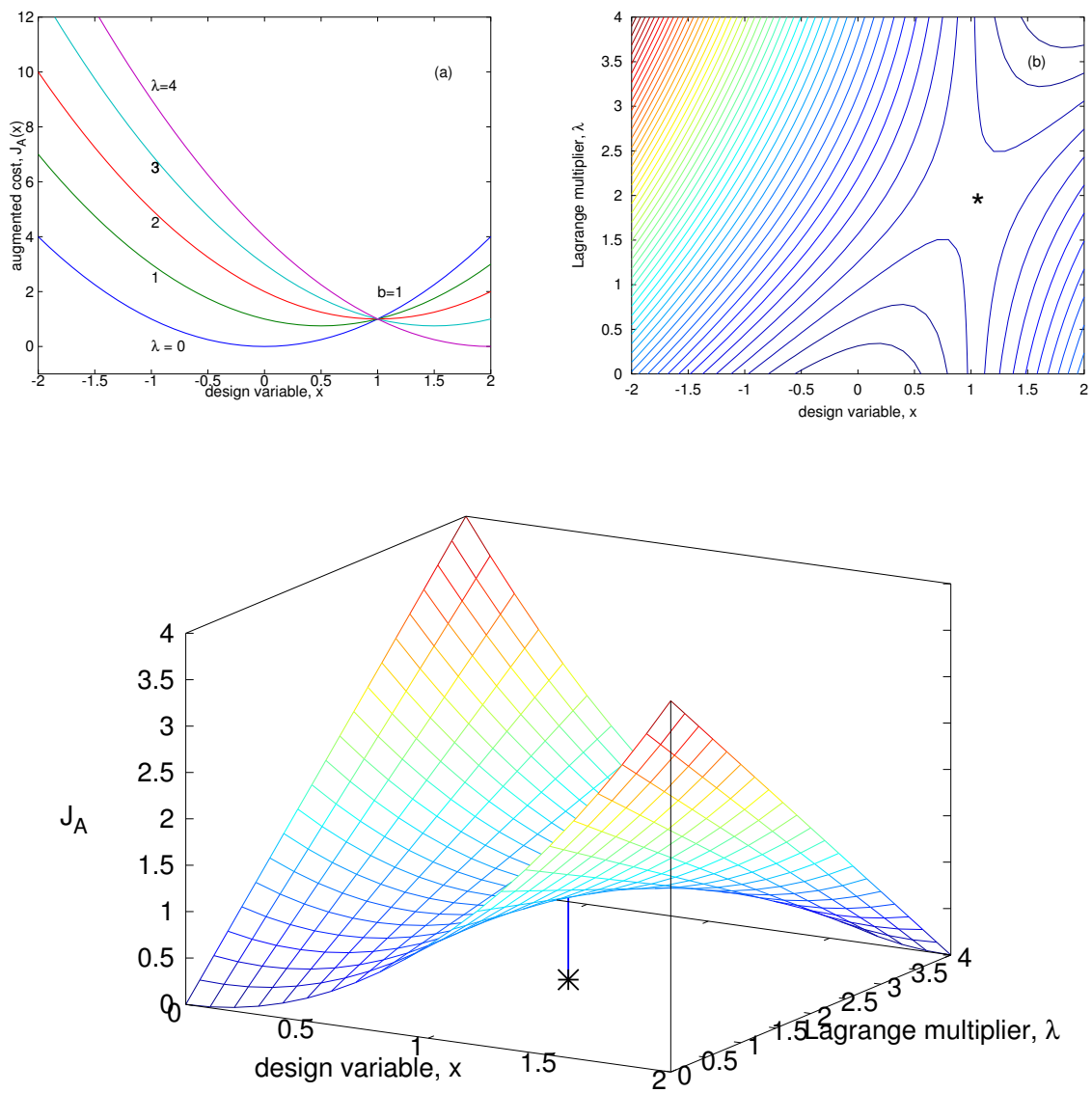

Figure 1. $J_A(x, \lambda) = \frac{1}{2}kx^2 - \lambda x + \lambda b$ for $b = 1$ and $k = 2$. $(x^*, \lambda^*) = (1, 2)$

Figure 1 shows that:

- $J_A(x, \lambda)$ is independent of λ at $x = b$,
- $J_A(x, \lambda)$ is minimized at $x^* = b$ for $\lambda^* = 2$,
- the surface $J_A(x, \lambda)$ is a saddle shape,
- the point $(x^*, \lambda^*) = (1, 2)$ is a saddle point,
- $J_A(x^*, \lambda) \leq J_A(x^*, \lambda^*) \leq J_A(x, \lambda^*)$,
- $\min_x J_A(x, \lambda^*) = \max_\lambda J_A(x^*, \lambda) = J_A(x^*, \lambda^*)$

Saddle points have no slope.

$$\left.\frac{\partial J_A(x,\lambda)}{\partial x}\right|_{\substack{x=x^*\\ \lambda=\lambda^*}} = 0 \tag{5a}$$

$$\left.\frac{\partial J_A(x,\lambda)}{\partial \lambda}\right|_{\substack{x=x^*\\ \lambda=\lambda^*}} = 0 \tag{5b}$$

$$\tag{5c}$$

For this problem,

$$\left.\frac{\partial J_A(x,\lambda)}{\partial x}\right|_{\substack{x=x^*\\ \lambda=\lambda^*}} = 0 \Rightarrow \quad kx^* - \lambda^* = 0 \quad \Rightarrow \lambda^* = kx^* \tag{6a}$$

$$\left.\frac{\partial J_A(x,\lambda)}{\partial \lambda}\right|_{\substack{x=x^*\\ \lambda=\lambda^*}} = 0 \Rightarrow \quad -x^* + b = 0 \quad \Rightarrow x^* = b \tag{6b}$$

This example has a physical interpretation. The objective function $J = \frac{1}{2}kx^2$ represents the potential energy in a spring. The minimum potential energy in a spring corresponds to a stretch of zero ($x^* = 0$). The constrained problem:

$$\min_x \ \frac{1}{2}kx^2 \qquad \text{such that} \qquad x \geq 1$$

means "minimize the potential energy in the spring such that the stretch in the spring is greater than or equal to 1." The solution to this problem is to set the stretch in the spring equal to the smallest allowable value ($x^* = 1$). The *force* applied to the spring in order to achieve this objective is $f = kx^*$. This force <u>*is*</u> the *Lagrange multiplier* for this problem, ($\lambda^* = kx^*$).

The Lagrange multiplier is the force required to enforce the constraint.

Case 2: $b = -1$

If $b = -1$ then the minimum of $\frac{1}{2}kx^2$ *is not* constrained by the inequality $x \geq b$. The derivation above would give $x^* = -1$, with $\lambda^* = -k$. The negative value of λ^* indicates that the constraint does not affect the optimal solution, and λ^* should therefore be set to zero. Setting $\lambda^* = 0$, $J_A(x, \lambda)$ is minimized at $x^* = 0$. Figure 2(a) plots $J_A(x, \lambda)$ for a few negative values of λ and Figure 2(b) plots contours of $J_A(x, \lambda)$.

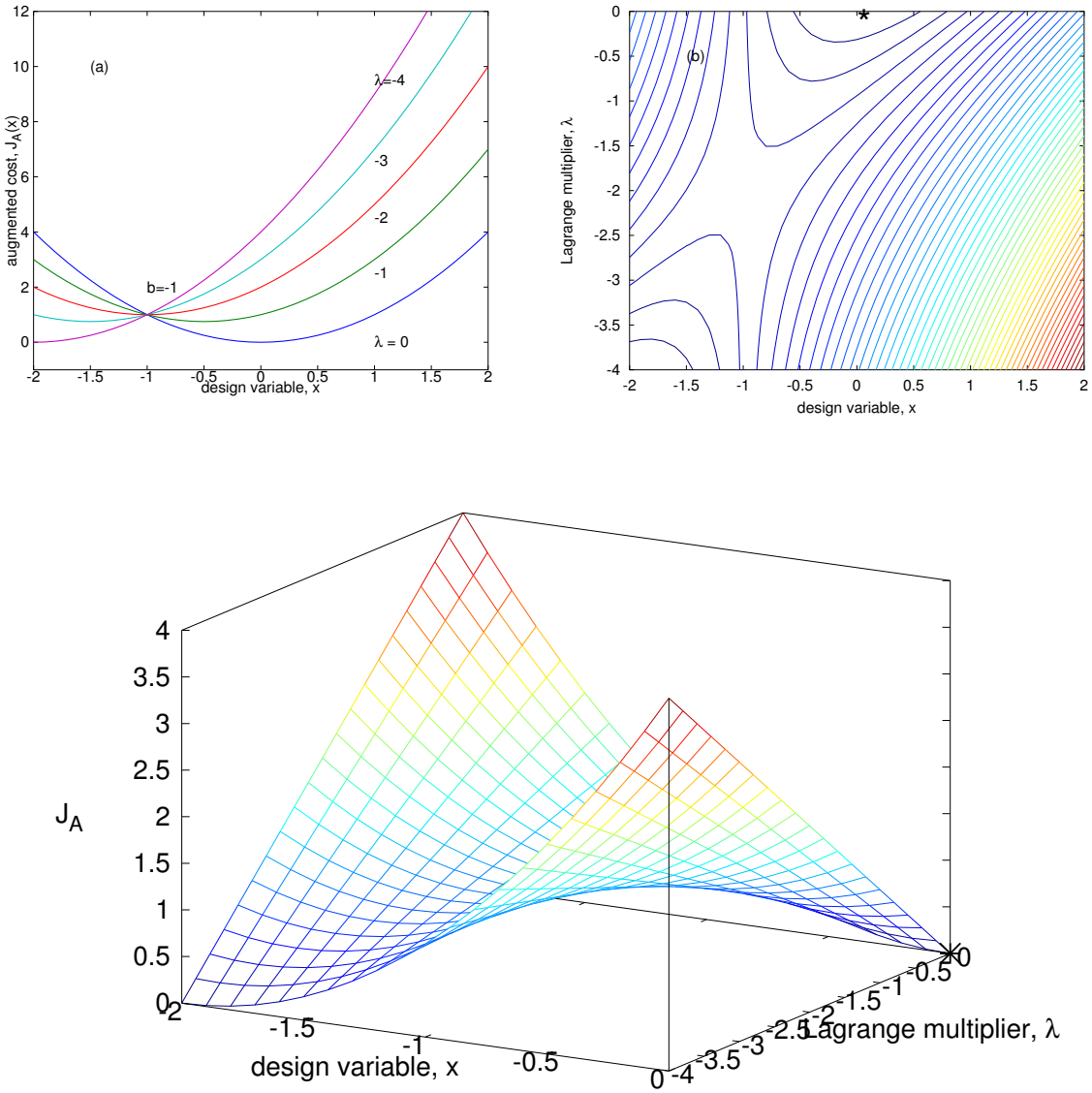

Figure 2. $J_A(x, \lambda) = \frac{1}{2}kx^2 - \lambda x + \lambda b$ for $b = -1$ and $k = 2$. $(x^*, \lambda^*) = (0, 0)$

Figure 2 shows that:

- $J_A(x, \lambda)$ is independent of λ at $x = b$,
- the saddle point of $J_A(x, \lambda)$ occurs at a negative value of λ, so $\partial J_A/\partial\lambda \neq 0$ for any $\lambda \geq 0$.
- The constraint $x \geq -1$ does not affect the solution, and is called a *non-binding* or an *inactive* constraint.
- The Lagrange multipliers associated with non-binding inequality constraints are negative.
- If a Lagrange multiplier corresponding to an inequality constraint has a negative value at the saddle point, it is set to zero, thereby removing the inactive constraint from the calculation of the augmented objective function.

Summary

In summary, if the inequality $g(x) \leq 0$ constrains the minimum of $f(x)$ then the optimum point of the augmented objective $J_A(x, \lambda) = f(x) + \lambda g(x)$ is minimized with respect to x and maximized with respect to λ. The optimization problem of equation (1) may be written in terms of the augmented objective function (equation (3)),

$$\max_{\lambda_1,\lambda_2,\cdots,\lambda_m} \min_{x_1,x_2,\cdots,x_n} J_A(x_1, x_2, \cdots, x_n,\ \lambda_1, \lambda_2, \cdots, \lambda_m) \qquad \text{such that} \qquad \lambda_j \geq 0\ \forall\ j \quad (7)$$

The necessary conditions for optimality are:

$$\left.\frac{\partial J_A}{\partial x_k}\right|_{\substack{x_i = x_i^* \\ \lambda_j = \lambda_j^*}} = 0 \tag{8a}$$

$$\left.\frac{\partial J_A}{\partial \lambda_k}\right|_{\substack{x_i = x_i^* \\ \lambda_j = \lambda_j^*}} = 0 \tag{8b}$$

$$\lambda_j \geq 0 \tag{8c}$$

These equations define the relations used to derive expressions for, or compute values of, the optimal design variables x_i^*. Lagrange multipliers for inequality constraints, $g_j(x_1, \cdots, x_n) \leq 0$, are non-negative. If an inequality $g_j(x_1, \cdots, x_n) \leq 0$ constrains the optimum point, the corresponding Lagrange multiplier, λ_j, is positive. If an inequality $g_j(x_1, \cdots, x_n) \leq 0$ does not constrain the optimum point, the corresponding Lagrange multiplier, λ_j, is set to zero.

Sensitivity to Changes in the Constraints and Redundant Constraints

Once a constrained optimization problem has been solved, it is sometimes useful to consider how changes in each constraint would affect the optimized cost. If the minimum of $f(\boldsymbol{x})$ (where $\boldsymbol{x} = (x_1, \cdots, x_n)$) is constrained by the inequality $g_j(\boldsymbol{x}) \leq 0$, then at the optimum point $\boldsymbol{x}^*$, $g_j(\boldsymbol{x}^*) = 0$ and $\lambda_j^* > 0$. Likewise, if another constraint $g_k(\boldsymbol{x}^*)$ does not constrain the minimum, then $g_k(\boldsymbol{x}^*) < 0$ and $\lambda_k^* = 0$. A small change in the j-th constraint, from g_j to $g_j + \delta g_j$, changes the constrained objective function from $f(\boldsymbol{x}^*)$ to $f(\boldsymbol{x}^*) + \lambda_j^* \delta g_j$. However, since the k-th constraint $g_k(\boldsymbol{x}^*)$ does not affect the solution, neither will a small change to g_k, $\lambda_k^* = 0$. In other words, a small change in the j-th constraint, δg_j, corresponds to a proportional change in the optimized cost, $\delta f(\boldsymbol{x}^*)$, and the constant of proportionality is the Lagrange multiplier of the optimized solution, λ_j^*.

$$\delta f(\boldsymbol{x}^*) = \sum_{j=1}^{m} \lambda_j^* \;\; \delta g_j \tag{9}$$

The value of the Lagrange multiplier is the sensitivity of the constrained objective to (small) changes in the constraint δg. If $\lambda_j > 0$ then the inequality $g_j(\boldsymbol{x}) \leq 0$ constrains the optimum point and a small increase of the constraint $g_j(\boldsymbol{x}^*)$ *increases* the cost. If $\lambda_j < 0$ then $g_j(\boldsymbol{x}^*) < 0$ and does not constrain the solution. A small change of this constraint should not affect the cost. So if a Lagrange multiplier associated with an inequality constraint $g_j(\boldsymbol{x}^*) \leq 0$ is computed as a negative value, it is subsequently set to zero. Such constraints are called *non-binding* or *inactive* constraints.

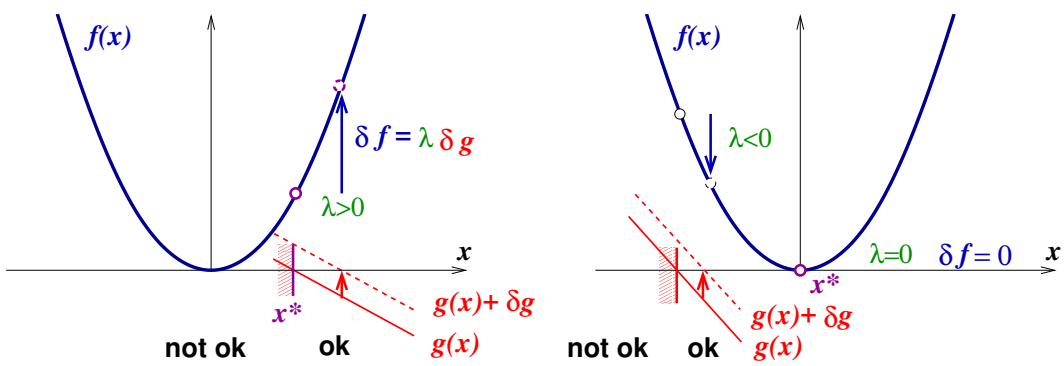

Figure 3. Left: The optimum point is constrained by the condition $g(x^*) \leq 0$. A small increase of the constraint from g to $g + \delta g$ *increases* the cost from $f(x^*)$ to $f(x^*) + \lambda \delta g$. ($\lambda > 0$) Right: The constraint is negative at the minimum of $f(x)$, so the optimum point is not constrained by the condition $g(x^*) \leq 0$. A small increase of the constraint function from g to $g + \delta g$ does not affect the optimum solution. If this constraint were an *equality* constraint, $g(x) = 0$, then the Lagrange multiplier would be negative and an increase of the constraint would *decrease* the cost from f to $f + \lambda \delta g$ ($\lambda < 0$).

The previous examples consider the minimization of a simple quadratic with a single inequality constraint. By expanding this example to two inequality constraints we can see again how Lagrange multipliers indicate whether or not the associated constraint bounds

the optimal solution. Consider the minimization problem

$$\min_x \frac{1}{2}kx^2 \quad \text{such that} \quad \begin{array}{l} x \geq b \\ x \geq c \end{array} \quad \text{where} \quad k > 0 \quad \text{and} \quad 0 < b < c \tag{10}$$

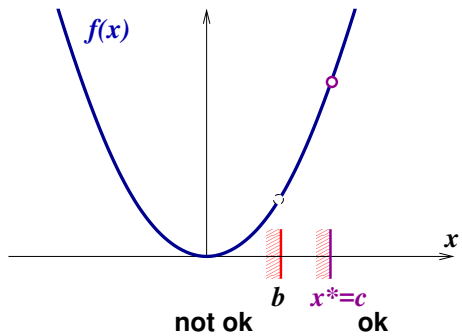

Figure 4. Minimization of $f(x) = \frac{1}{2}kx^2$ such that $x \geq b$ *and* $x \geq c$ with $0 < b < c$. Feasible solutions for x are greater-or-equal than both b and c.

Proceeding with the method of Lagrange multipliers

$$J_A(x_1\ \lambda_1, \lambda_2) = \frac{1}{2}kx^2 + \lambda_1(b - x) + \lambda_2(c - x) \tag{11}$$

Where the constraints $(b - x)$ and $(c - x)$ are satisfied if they are *negative-valued*. The derivatives of the augmented objective function are

$$\left.\frac{\partial J_A(x, \lambda_1, \lambda_2)}{\partial x}\right|_{\substack{x = x^* \\ \lambda_1 = \lambda_1^* \\ \lambda_2 = \lambda_2^*}} = 0 \Rightarrow \quad kx^* - \lambda_1^* - \lambda_2^* = 0 \tag{12a}$$

$$\left.\frac{\partial J_A(x, \lambda_1, \lambda_2)}{\partial \lambda_1}\right|_{\substack{x = x^* \\ \lambda_1 = \lambda_1^* \\ \lambda_2 = \lambda_2^*}} = 0 \Rightarrow \quad -x^* + b = 0 \tag{12b}$$

$$\left.\frac{\partial J_A(x, \lambda_1, \lambda_2)}{\partial \lambda_2}\right|_{\substack{x = x^* \\ \lambda_1 = \lambda_1^* \\ \lambda_2 = \lambda_2^*}} = 0 \Rightarrow \quad -x^* + c = 0 \tag{12c}$$

These three equations are linear in x^*, λ_1^*, and λ_2^*.

$$\begin{bmatrix} k & -1 & -1 \\ -1 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x^* \\ \lambda_1^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} 0 \\ -b \\ -c \end{bmatrix} \tag{13}$$

Clearly, if $b \neq c$ then x can not equal both b and c, and equations (12b) and (12c) can not both be satisfied. In such cases, we proceed by selecting a subset of the constraints, and evaluating the resulting solutions. So doing, the solution $x^* = b$ minimizes $\frac{1}{2}kx^2$ such that $x \geq b$ with a Lagrange multiplier $\lambda_1 = kb$. But $x^* = b$ does not satisfy the inequality $x \leq c$. The solution $x^* = c$ minimizes $\frac{1}{2}kx^2$. such that both inequalities hold and with with a Lagrange multiplier $\lambda_2 = kc$. Note that for this quadratic objective, the constraint with the larger Lagrange multiplier is the active constraint and that the other constraint is non-binding.

Multivariable Quadratic Programming

Quadratic optimization problems with a single design variable and a single linear inequality constraint are easy enough. In problems with many design variables ($n \gg 1$) and many inequality constraints ($m \gg 1$), determining which inequality constraints are enforced at the optimum point can be difficult. Numerical methods used in solving such problems involve iterative trial-and-error approaches to find the set of "active" inequality constraints.

Consider a third example with $n = 2$ and $m = 2$:

$$\min_{x_1,x_2} \; x_1^2 + 0.5x_1 + 3x_1x_2 + 5x_2^2 \qquad \text{such that} \qquad \begin{array}{rcrcl} g_1 & = & 3x_1 + 2x_2 + 2 & \leq & 0 \\ g_2 & = & 15x_1 - 3x_2 - 1 & \leq & 0 \end{array} \tag{14}$$

This example also has a quadratic objective function and inequality constraints that are linear in the design variables. Contours of the objective function and the two inequality constraints are plotted in Figure 5. The feasible region of these two inequality constraints is to the left of the lines in the figure and are labeled as "g_1 ok" and "g_2 ok". This figure shows that the inequality $g_1(x_1, x_2)$ constrains the solution and the inequality $g_2(x_1, x_2)$ does not. This is visible in Figure 5 with $n = 2$, but for more complicated problems it may not be immediately clear which inequality constraints are "active."

Using the method of Lagrange multipliers, the augmented objective function is

$$J_A(x_1, x_2,\ \lambda_1, \lambda_2) = x_1^2 + 0.5x_1 + 3x_1x_2 + 5x_2^2 + \lambda_1(3x_1 + 2x_2 + 2) + \lambda_2(15x_1 - 3x_2 - 1) \tag{15}$$

Unlike the first examples with $n = 1$ and $m = 1$, we cannot plot contours of $J_A(x_1, x_2, \lambda_1, \lambda_2)$ since this would be a plot in four-dimensional space. Nonetheless, the same optimality conditions hold.

$$\min_{x_1} J_A \Rightarrow \left.\frac{\partial J_A}{\partial x_1}\right|_{x_1^*,x_2^*,\lambda_1^*,\lambda_2^*} = 0 \Rightarrow \quad 2x_1^* + 0.5 + 3x_2^* + 3\lambda_1^* + 15\lambda_2^* = 0 \tag{16a}$$

$$\min_{x_2} J_A \Rightarrow \left.\frac{\partial J_A}{\partial x_2}\right|_{x_1^*,x_2^*,\lambda_1^*,\lambda_2^*} = 0 \Rightarrow \quad 3x_1^* + 10x_2^* + 2\lambda_1^* - 3\lambda_2^* = 0 \tag{16b}$$

$$\max_{\lambda_1} J_A \Rightarrow \left.\frac{\partial J_A}{\partial \lambda_1}\right|_{x_1^*,x_2^*,\lambda_1^*,\lambda_2^*} = 0 \Rightarrow \quad 3x_1^* + 2x_2^* + 2 = 0 \tag{16c}$$

$$\max_{\lambda_2} J_A \Rightarrow \left.\frac{\partial J_A}{\partial \lambda_2}\right|_{x_1^*,x_2^*,\lambda_1^*,\lambda_2^*} = 0 \Rightarrow \quad 15x_1^* - 3x_2^* - 1 = 0 \tag{16d}$$

If the objective function is quadratic in the design variables and the constraints are linear in the design variables, the optimality conditions are simply a set of linear equations in the design variables and the Lagrange multipliers. In this example the optimality conditions are expressed as four linear equations with four unknowns. In general we may not know which inequality constraints are active. If there are only a few constraint equations it's not too hard to try all combinations of any number of constraints, fixing the Lagrange multipliers for the other inequality constraints equal to zero.

Let's try this now!

- First, let's find the unconstrained minimum by assuming neither constraint $g_1(x_1, x_2)$ or $g_2(x_1, x_2)$ is active, $\lambda_1^* = 0$, $\lambda_2^* = 0$, and

$$\begin{bmatrix} 2 & 3 \\ 3 & 10 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0 \end{bmatrix} \qquad \Rightarrow \qquad \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \begin{bmatrix} -0.45 \\ 0.14 \end{bmatrix}, \tag{17}$$

  which is the unconstrained minimum shown in Figure 5 as a "$*$". Plugging this solution into the constraint equations gives $g_1(x_1^*, x_2^*) = 0.93$ and $g_2(x_1^*, x_2^*) = -8.17$, so the unconstrained minimum is not feasible with respect to constraint g_1, since $g_1(-0.45, 0.14) > 0$.

- Next, assuming both constraints $g_1(x_1, x_2)$ and $g_2(x_1, x_2)$ are active, optimal values for x_1^*, x_2^*, λ_1^*, and λ_2^* are sought, and all four equations must be solved together.

$$\begin{bmatrix} 2 & 3 & 3 & 15 \\ 3 & 10 & 2 & -3 \\ 3 & 2 & 0 & 0 \\ 15 & -3 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \\ \lambda_1^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0 \\ -2 \\ 1 \end{bmatrix} \qquad \Rightarrow \qquad \begin{bmatrix} x_1^* \\ x_2^* \\ \lambda_1^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} -0.10 \\ -0.85 \\ 3.55 \\ -0.56 \end{bmatrix}, \tag{18}$$

  which is the constrained minimum shown in Figure 5 as a "o" at the intersection of the g_1 line with the g_2 line in Figure 5. Note that $\lambda_2^* < 0$ indicating that constraint g_2 is not active; $g_2(-0.10, 0.85) = 0$ (*ok*). Enforcing the constraint g_2 needlessly compromises the optimality of this solution. So, while this solution is feasible (both g_1 and g_2 evaluate to zero), the solution could be improved by letting go of the g_2 constraint and moving along the g_1 line.

- So, assuming only constraint g_1 is active, g_2 is not active, $\lambda_2^* = 0$, and

$$\begin{bmatrix} 2 & 3 & 3 \\ 3 & 10 & 2 \\ 3 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \\ \lambda_1^* \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0 \\ -2 \end{bmatrix} \qquad \Rightarrow \qquad \begin{bmatrix} x_1^* \\ x_2^* \\ \lambda_1^* \end{bmatrix} = \begin{bmatrix} -0.81 \\ 0.21 \\ 0.16 \end{bmatrix}, \tag{19}$$

  which is the constrained minimum as a "o" on the g_1 line in Figure 5. Note that $\lambda_1^* > 0$, which indicates that this constraint is active. Plugging x_1^* and x_2^* into $g_2(x_1, x_2)$ gives a value of -13.78, so this constrained minimum is feasible with respect to both constraints. This is the solution we're looking for.

- As a final check, assuming only constraint g_2 is active, g_1 is not active, $\lambda_1^* = 0$, and

$$\begin{bmatrix} 2 & 3 & 15 \\ 3 & 10 & -3 \\ 15 & -3 & 0 \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} -0.5 \\ 0 \\ 1 \end{bmatrix} \qquad \Rightarrow \qquad \begin{bmatrix} x_1^* \\ x_2^* \\ \lambda_2^* \end{bmatrix} = \begin{bmatrix} 0.06 \\ -0.03 \\ -0.04 \end{bmatrix}, \tag{20}$$

  which is the constrained minimum shown in as a "o" on the g_2 line in Figure 5. Note that $\lambda_2^* < 0$, which indicates that this constraint is not active, contradicting our assumption. Further, plugging x_1^* and x_2^* into $g_1(x_1, x_2)$ gives a value of $+2.24$, so this constrained minimum is not feasible with respect to g_1, since $g_1(0.06, -0.03) > 0$.

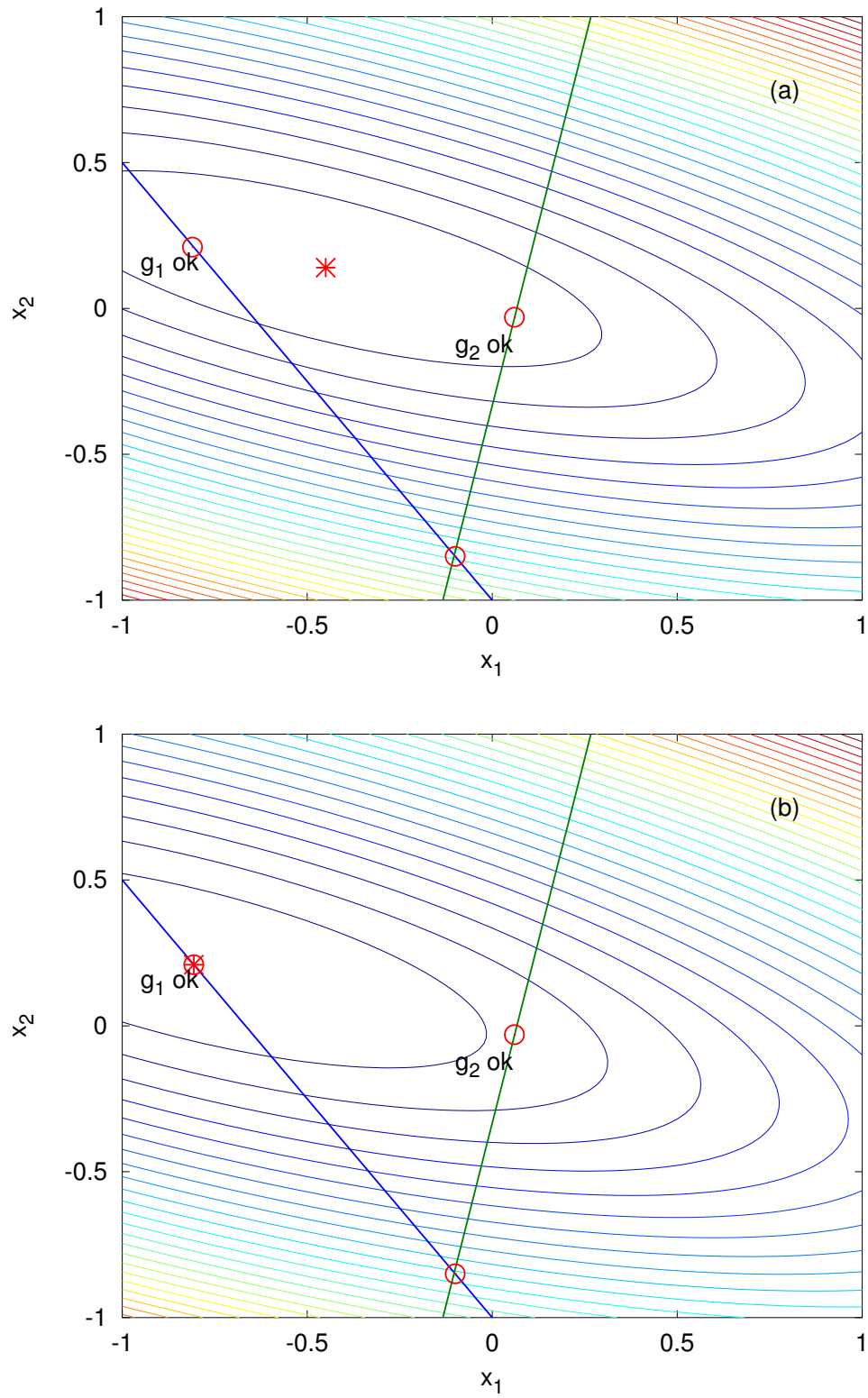

Figure 5. Contours of the objective function and the constraint equations for the example of equation (14). (a): $J = f(x_1, x_2)$; (b): $J_A = f(x_1, x_2) + \lambda_1^* g_1(x_1, x_2)$. Note that the the contours of J_A are shifted so that the minimum of J_A is at the optimal point along the g_1 line.

Matrix Relations for the Minimum of a Quadratic with Equality Constraints

The previous example of the minimization of a quadratic in x_1 and x_2, subject to two in linear equality constraints equations (14), may be written, generally as

$$\min_{x_1,x_2} \frac{1}{2}H_{11}x_1^2 + H_{12}x_1x_2 + \frac{1}{2}H_{22}x_2^2 + c_1x_1 + c_2x_2 + d \quad \text{such that} \quad \begin{array}{c} A_{11}x_1 + A_{12}x_2 - b_1 \leq 0 \\ A_{21}x_1 + A_{22}x_2 - b_2 \leq 0 \end{array} \tag{21}$$

or, even more generally, for any number (n) of design variables, as

$$\min_{x_1,x_2,\cdots,x_n} \left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_iH_{ij}x_j + \sum_{i=1}^{n} c_ix_i + d\right] \quad \text{such that} \quad \begin{array}{c} \sum_{j=1}^{n} A_{1j}x_j - b_1 \leq 0 \\ \sum_{j=1}^{n} A_{2j}x_j - b_2 \leq 0 \\ \vdots \\ \sum_{j=1}^{n} A_{mj}x_j - b_m \leq 0 \end{array} \tag{22}$$

where $H_{ij} = H_{ji}$ and $\sum_i \sum_j x_iH_{ij}x_j > 0$ for any set of design variables. In matrix-vector notation, equation (22) is written

$$\min_{\boldsymbol{x}} \ \frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d \qquad \text{such that} \qquad \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \leq \boldsymbol{0} \tag{23}$$

The augmented optimization problem is

$$\max_{\lambda_1,\cdots,\lambda_m} \min_{x_1,\cdots,x_n} \left[\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_iH_{ij}x_j + \sum_{i=1}^{n} c_ix_i + d + \sum_{k=1}^{m}\lambda_k\left(\sum_{j=1}^{n} A_{kj}x_j - b_k\right)\right] , \tag{24}$$

such that $\lambda_k \geq 0$. Or, in matrix-vector notation,

$$\max_{\boldsymbol{\lambda}} \min_{\boldsymbol{x}} \ \left[\frac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c}^\mathsf{T}\boldsymbol{x} + d + \boldsymbol{\lambda}^\mathsf{T}\left(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\right)\right] \qquad \text{such that} \qquad \boldsymbol{\lambda} \geq \boldsymbol{0} \ . \tag{25}$$

Assuming all constraints are "active" the necessary conditions for optimality are

$$\frac{\partial J_A(\boldsymbol{x},\boldsymbol{\lambda})}{\partial \boldsymbol{x}} = \boldsymbol{0}^\mathsf{T} \qquad \text{and} \qquad \frac{\partial J_A(\boldsymbol{x},\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \boldsymbol{0}^\mathsf{T} \tag{26}$$

Which, for this very general system result in

$$\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c} + \boldsymbol{A}^\mathsf{T}\boldsymbol{\lambda} = \boldsymbol{0} \qquad \text{and} \qquad \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0} \tag{27}$$

These equations can be written compactly as a single matrix equation

$$\begin{bmatrix} \boldsymbol{H} & \boldsymbol{A}^\mathsf{T} \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{c} \\ \boldsymbol{b} \end{bmatrix} \tag{28}$$

Equation (28) is called the Karush Kuhn Tucker (KKT) equation.

Comparing equation (14) to equation (22) and comparing equations (18), (19), and (20) to equation (28), you may begin to convince yourself that the KKT system gives a general solution for the minimum of a quadratic with linear equality constraints. It is very useful.

Primal and Dual

If it is possible to solve for the design variables $\boldsymbol{x}$ in terms of the Lagrange multipliers $\boldsymbol{\lambda}$, then the design variables can be eliminated from the problem and the optimization is simply a maximization over the set of Lagrange multipliers.

For this kind of problem, the condition

$$\frac{\partial J_A(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial \boldsymbol{x}} = \mathbf{0}^\mathsf{T} \tag{29}$$

results in $\boldsymbol{H}\boldsymbol{x} + \boldsymbol{c} + \boldsymbol{A}^\mathsf{T}\boldsymbol{\lambda} = \mathbf{0}$ from which the design variables $\boldsymbol{x}$ may be solved in terms of $\boldsymbol{\lambda}$

$$\boldsymbol{x} = -\boldsymbol{H}^{-1}(\boldsymbol{A}^\mathsf{T}\boldsymbol{\lambda} + \boldsymbol{c}). \tag{30}$$

Substituting this solution for $\boldsymbol{x}$ in terms of $\boldsymbol{\lambda}$ into (25) eliminates $\boldsymbol{x}$ from equation (25) and results in a really long expression for the augmented objective in terms of $\boldsymbol{\lambda}$ only

$$\max_{\boldsymbol{\lambda}} \left[\frac{1}{2}\left(-(\boldsymbol{\lambda}^\mathsf{T}\boldsymbol{A} + \boldsymbol{c}^\mathsf{T})\boldsymbol{H}^{-1}\right)\boldsymbol{H}\left(-\boldsymbol{H}^{-1}(\boldsymbol{A}^\mathsf{T}\boldsymbol{\lambda} + \boldsymbol{c})\right) - \boldsymbol{c}^\mathsf{T}\boldsymbol{H}^{-1}(\boldsymbol{A}^\mathsf{T}\boldsymbol{\lambda} + \boldsymbol{c}) + d + \boldsymbol{\lambda}^\mathsf{T}\left(\boldsymbol{A}\left(-\boldsymbol{H}^{-1}(\boldsymbol{A}^\mathsf{T}\boldsymbol{\lambda} + \boldsymbol{c})\right) - \boldsymbol{b}\right)\right]$$

... and after a little algebra involving some convenient cancellations and combinations ... we obtain the *dual* problem.

$$\max_{\boldsymbol{\lambda}} \left[-\frac{1}{2}\boldsymbol{\lambda}^\mathsf{T}\boldsymbol{A}\boldsymbol{H}^{-1}\boldsymbol{A}^\mathsf{T}\boldsymbol{\lambda} - (\boldsymbol{c}^\mathsf{T}\boldsymbol{H}^{-1}\boldsymbol{A}^\mathsf{T} + \boldsymbol{b}^\mathsf{T})\boldsymbol{\lambda} - \frac{1}{2}\boldsymbol{c}^\mathsf{T}\boldsymbol{H}^{-1}\boldsymbol{c} + d\right] \quad \text{such that} \quad \boldsymbol{\lambda} \geq \mathbf{0} \tag{31}$$

Equation (25) is called the *primal* quadratic programming problem and equation (31) is called the *dual* quadratic programming problem. The primal problem has $n+m$ unknown variables ($\boldsymbol{x}$ and $\boldsymbol{\lambda}$) whereas the dual problem has only m unknown variables ($\boldsymbol{\lambda}$) and is therefore easier to solve.

Note that if quadratic term in the primal form (quadratic in the primal variable $\boldsymbol{x}$) is positive, then the quadratic term in the dual form (quadratic in the dual variable $\boldsymbol{\lambda}$) is negative. So, again, it makes sense to *minimize* over the primal variable $\boldsymbol{x}$ and to *maximize* over the dual variable $\boldsymbol{\lambda}$.

An Active Set Algorithm for Quadratic Programming

As was shown in the numerical example, in problems with multiple inequality constraints it is often hard to know which of the inequality constraints are active. A trial-and-error approach may be attempted to find the set of active inequality constriants, for which all the Lagrange multipliers are positive-valued. Anticipating nonlinear problems, we will assume we know that $\boldsymbol{x}$ is a good but potentially infeasible solution to the quadratic program. We wish to update $\boldsymbol{x}$ to a better solution by adding the update vector $\boldsymbol{h}$, so that $\boldsymbol{x}+\boldsymbol{h}$ solves the quadratic program. To find the "optimal" update vector, we can express the quadratic in terms of the known current parameter values $\boldsymbol{x}$ and the unknown update vector $\boldsymbol{h}$,

$$\max_{\boldsymbol{\lambda}} \min_{\boldsymbol{h}} \left[\frac{1}{2}(\boldsymbol{x}+\boldsymbol{h})^{\mathsf{T}}\boldsymbol{H}(\boldsymbol{x}+\boldsymbol{h}) + \boldsymbol{c}^{\mathsf{T}}(\boldsymbol{x}+\boldsymbol{h}) + d + \boldsymbol{\lambda}^{\mathsf{T}}\left(\boldsymbol{A}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{b}\right)\right] \quad \text{s.t.} \quad \boldsymbol{\lambda} \geq \boldsymbol{0} \tag{32}$$

Expanding equation (32) and deriving the necessary conditions for optimality of the step $\boldsymbol{h}$,

$$\frac{\partial J_A(\boldsymbol{h},\boldsymbol{\lambda})}{\partial \boldsymbol{h}} = \boldsymbol{0}^{\mathsf{T}} \qquad \text{and} \qquad \frac{\partial J_A(\boldsymbol{h},\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = \boldsymbol{0}^{\mathsf{T}} \tag{33}$$

leads to

$$\boldsymbol{H}(\boldsymbol{x}+\boldsymbol{h}) + \boldsymbol{c} + \boldsymbol{A}^{\mathsf{T}}\boldsymbol{\lambda} = \boldsymbol{0} \qquad \text{and} \qquad \boldsymbol{A}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{b} = \boldsymbol{0} \tag{34}$$

which may be writtin in a KKT form

$$\begin{bmatrix} \boldsymbol{H} & \boldsymbol{A}^{\mathsf{T}} \\ \boldsymbol{A} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}\boldsymbol{x} - \boldsymbol{c} \\ -\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \end{bmatrix} \tag{35}$$

But the question remains, "Which constraint equations are active?" In the following the constraint equations in the *active set* are denoted

$$\boldsymbol{A}_{\mathsf{a}}\boldsymbol{x} - \boldsymbol{b}_{\mathsf{a}} = \boldsymbol{0} \tag{36}$$

with Lagrange multipliers $\boldsymbol{\lambda}_{\mathsf{a}}$ associated with the constraint equations in the active set. In this active set algorithm, we may start by assuming no constraints are active, so that $\boldsymbol{A}_{\mathsf{a}}^{(0)} = \boldsymbol{A}$ and $\boldsymbol{b}_{\mathsf{a}}^{(0)} = \boldsymbol{b}$.

1. In iteration (k), solve

$$\begin{bmatrix} \boldsymbol{H} & \boldsymbol{A}_{\mathsf{a}}^{(k)\mathsf{T}} \\ \boldsymbol{A}_{\mathsf{a}}^{(k)} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{\lambda}_{\mathsf{a}}^{(k)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}\boldsymbol{x} - \boldsymbol{c} \\ -\boldsymbol{A}_{\mathsf{a}}^{(k)}\boldsymbol{x} + \boldsymbol{b}_{\mathsf{a}}^{(k)} \end{bmatrix}$$

   for $\boldsymbol{h}$ and $\boldsymbol{\lambda}_{\mathsf{a}}^{(k)}$.

2. Evaluate the full set of constraint equations $\boldsymbol{g} = \boldsymbol{A}(\boldsymbol{x}+\boldsymbol{h}) - \boldsymbol{b}$.

3. In $\boldsymbol{A}_{\mathsf{a}}^{(k+1)}$ and $\boldsymbol{b}^{(k+1)}$ include the rows of $\boldsymbol{A}$ and $\boldsymbol{b}$ corresponding to the rows of $\boldsymbol{g}$ that are positive.

4. In $\boldsymbol{A}_{\mathsf{a}}^{(k+1)}$ and $\boldsymbol{b}^{(k+1)}$ exclude the rows of $\boldsymbol{A}$ and $\boldsymbol{b}$ corresponding to the rows of $\boldsymbol{\lambda}_{\mathsf{a}}^{(k)}$ that are negative.

5. Assign elements of $\boldsymbol{\lambda}_{\mathsf{a}}^{(k)}$ to the active indices of $\boldsymbol{\lambda}$.

6. set $k = k+1$ and return to step 1.