Solutions Chapter 3

SECTION 3.2

3.2.6 (www)

Assume that the matrix

$$J = \begin{pmatrix} \nabla^2_{xx} L(x^*, \lambda^*) & \nabla h(x^*) \\ \nabla h(x^*)' & 0 \end{pmatrix}$$

is invertible, but the sufficiency conditions do not hold for x^* and λ^*. Since x^* and λ^* satisfy the first and the second order necessary conditions of Prop. 3.2.1, this implies that there is a vector $\bar{y} \neq 0$ such that $\nabla h(x^*)'\bar{y} = 0$ and $\bar{y}'\nabla^2_{xx} L(x^*, \lambda^*)\bar{y} = 0$. Hence, $\bar{y}$ minimizes the quadratic function $y'\nabla^2_{xx} L(x^*, \lambda^*)y$ over all y with $\nabla h(x^*)'y = 0$. Thus $\nabla^2_{xx} L(x^*, \lambda^*)\bar{y} = 0$, and we have

$$\begin{pmatrix} \nabla^2_{xx} L(x^*, \lambda^*) & \nabla h(x^*) \\ \nabla h(x^*)' & 0 \end{pmatrix} \begin{pmatrix} \bar{y} \\ 0 \end{pmatrix} = 0,$$

which contradict the invertibility of J.

For the reverse assertion, assume that x^* and λ^* satisfy the second order sufficiency conditions of Prop. 3.2.1. Let $\bar{y} \in \Re^n$ and $\bar{z} \in \Re^m$ be vectors such that

$$J \begin{pmatrix} \bar{y} \\ \bar{z} \end{pmatrix} = 0.$$

Consequently

$$\nabla^2_{xx} L(x^*, \lambda^*)\bar{y} + \nabla h(x^*)\bar{z} = 0, \tag{1}$$

$$\nabla h(x^*)'\bar{y} = 0. \tag{2}$$

Pre-multiplying Eq. (1) by $\bar{y}$ and using Eq. (2), we obtain

$$\bar{y}\nabla^2_{xx} L(x^*, \lambda^*)\bar{y} = 0.$$

In view of Eq. (2), it follows that $\bar{y} = 0$, for otherwise the second order sufficiency condition would be violated. Then Eq. (1) yields $\nabla h(x^*)\bar{z} = 0$. Since x^* is a regular point, we must have $\bar{z} = 0$. Hence, J is invertible.

3.2.7 (www)

We have

$$\nabla^2 p(u) = -\nabla\lambda(u).$$

To calculate $\nabla\lambda(u)$, we differentiate the relation

$$\nabla f\big(x(u)\big) + \nabla h\big(x(u)\big)\lambda(u) = 0.$$

We have

$$\nabla x(u)\nabla^2_{xx}L\big(x(u),\lambda(u)\big) + \nabla\lambda(u)\nabla h\big(x(u)\big)' = 0.$$

We also have $\nabla x(u)\nabla h\big(x(u)\big) = I$, from which we obtain for all $c \in \Re$

$$c\nabla x(u)\nabla h\big(x(u)\big)\nabla h\big(x(u)\big)' = c\nabla h\big(x(u)\big)'.$$

By adding the last two equations, we see that

$$\nabla x(u)\big(\nabla^2_{xx}L\big(x(u),\lambda(u)\big) + c\nabla h\big(x(u)\big)\nabla h\big(x(u)\big)'\big) + \big(\nabla\lambda(u) - cI\big)h\big(x(u)\big)' = 0.$$

From this, we obtain, for every c for which the inverse below exists,

$$\nabla x(u) + \big(\nabla\lambda(u) - cI\big)h\big(x(u)\big)'\Big(\nabla^2_{xx}L\big(x(u),\lambda(u)\big) + c\nabla h\big(x(u)\big)\nabla h\big(x(u)\big)'\Big)^{-1} = 0.$$

Multiplying with $\nabla h\big(x(u)\big)$ and using the equations $\nabla x(u)\nabla h\big(x(u)\big) = I$ and $\nabla^2 p(u) = -\nabla\lambda(u)$, we see that

$$\nabla^2 p(u) = \Big(\nabla h\big(x(u)\big)'\big(\nabla^2_{xx}L\big(x(u),\lambda(u)\big) + c\nabla h\big(x(u)\big)\nabla h\big(x(u)\big)'\big)^{-1}\nabla h\big(x(u)\big)\Big)^{-1} - cI.$$

SECTION 3.3

3.3.5 (www)

(a) Let $d \in \overline{F}(x^*)$ be arbitrary. Then there exists a sequence $\{d^k\} \subseteq F(x^*)$ such that $d^k \to d$. For each d^k, we have

$$\nabla f(x^*)'d^k = \lim_{\alpha \to 0} \frac{f(x^* + \alpha d^k) - f(x^*)}{\alpha}.$$

Since x^* is a constrained local minimum, we have $\frac{f(x^*+\alpha d^k)-f(x^*)}{\alpha} \geq 0$ for all sufficiently small α (for which $x^* + \alpha d^k$ is feasible), and thus $\nabla f(x^*)'d^k \geq 0$. Hence

$$\nabla f(x^*)'d = \lim_{k \to \infty} \nabla f(x^*)'d^k \geq 0$$

as desired.

(b) If x^* is a constrained local minimum, we have from part (a)

$$\nabla f(x^*)'d \geq 0 \quad \forall\, d \text{ with } \nabla g_j(x^*)'d \leq 0, \quad \forall\, j \in A(x^*).$$

According to Farkas' lemma, this is true if and only if there exists μ^* such that

$$-\nabla f(x^*) = \sum_{j \in A(x^*)} \mu_j^* \nabla g_j(x^*), \quad \mu_j^* \geq 0.$$

Setting $\mu_j^* = 0$ for $j \notin A(x^*)$, we have the desired result.

(c) We want to show that $\overline{F}(x^*) = V(x^*)$, where $V(x^*)$ is the cone of first order feasible variations given by

$$V(x^*) = \big\{d \mid \nabla g_j(x^*)'d \leq 0, \forall\, j \in A(x^*)\big\}.$$

First, let's show that under any of the conditions (1)–(4), we have $F(x^*) \subseteq V(x^*)$. By Mean Value Theorem, for each $j \in A(x^*)$ and for any $d \in F(x^*)$ there is some $\epsilon \in [0,1]$ such that

$$g_j(x^* + \alpha d) = g_j(x^*) + \alpha \nabla g_j(x^* + \epsilon \alpha d)'d.$$

Because $g_j(x^* + \alpha d) \leq 0$ for all $\alpha \in [0, \bar{\alpha}]$ and $g_j(x^*) = 0$ for all $j \in A(x^*)$, we obtain for all $j \in A(x^*)$

$$\lim_{\alpha \to 0} \nabla g_j(x^* + \epsilon \alpha d)'d \leq 0,$$

which by continuity of each ∇g_j implies that

$$\nabla g_j(x^*)'d \leq 0, \qquad \forall\, j \in A(x^*),$$

so that $d \in V(x^*)$. Therefore $F(x^*) \subseteq V(x^*)$ and $\overline{F}(x^*) \subseteq V(x^*)$ [because $V(x^*)$ is closed].

Now we need to show that $V(x^*) \subseteq \overline{F}(x^*)$ for each of the parts (1) through (4).

(1) Let $g_j(x) = b_j'x + c_j$ for all j, where b_j are vectors and c_j are scalars. Let $d \in V(x^*)$. We have

$$g_j(x^* + \alpha d) = b_j'(x^* + \alpha d) + c_j = g_j(x^*) + \alpha b_j' d.$$

If $j \in A(x^*)$, then by the definition of $V(x^*)$ we have $b_j'd = \nabla g_j(x^*)'d \leq 0$, so that $g_j(x^* + \alpha d) \leq g_j(x^*) = 0$ for all $\alpha > 0$. If $j \notin A(x^*)$ and $b_j'd \leq 0$, then $g_j(x^* + \alpha d) \leq g_j(x^*) < 0$ for any $\alpha > 0$ [because this constraint is not tight at x^*]. If $j \notin A(x^*)$ and $b_j'd > 0$, then $g_j(x^* + \alpha d) \leq 0$ for all $\alpha \leq \bar{\alpha}_j$, where $\bar{\alpha}_j = -g_j(x^*)/(a_j'd)$ [here we use $g_j(x^*) < 0$]. Therefore we have $g_j(x^* + \alpha d) \leq 0$ for all j and all $\alpha \leq \bar{\alpha}$, where

$$\bar{\alpha} = \min\bigl\{\bar{\alpha}_j \mid j \notin A(x^*),\ b_j'd > 0\bigr\}.$$

Thus $d \in F(x^*)$ and consequently $V(x^*) \subseteq \overline{F}(x^*)$ [since $V(x^*)$ is closed].

(2) Let $d \in V(x^*)$ and let $\bar{d}$ be such that

$$\nabla g_j(x^*)'\bar{d} < 0, \qquad \forall\ j \in A(x^*).$$

Define $d_\gamma = \gamma \bar{d} + (1 - \gamma)d$. By using the Mean Value Theorem, for each j there is some $\epsilon \in [0, 1]$ such that

$$\begin{aligned} g_j(x^* + \alpha d_\gamma) &= g_j(x^*) + \alpha \nabla g_j(x^* + \epsilon \alpha d_\gamma)'d_\gamma \\ &= g_j(x^*) + \alpha\gamma \nabla g_j(x^* + \epsilon\alpha d_\gamma)'\bar{d} + \alpha(1-\gamma)\nabla g_j(x^* + \epsilon \alpha d_\gamma)'d. \end{aligned}$$

Let γ be fixed. If $j \notin A(x^*)$, then by using the fact $g_j(x^*) < 0$ it can be seen that for all sufficiently small α we have

$$g_j(x^* + \alpha d_\gamma) \leq 0, \qquad \forall\ j \notin A(x^*).$$

If $j \in A(x^*)$, then by continuity of ∇g_j we have for all sufficiently small α

$$\nabla g_j(x^* + \epsilon \alpha d_\gamma)'\bar{d} \leq 0.$$

This combined with the fact $d \in V(x^*)$ implies that for all sufficiently small α

$$g_j(x^* + \alpha d_\gamma) \leq 0, \qquad \forall\ j \in A(x^*).$$

Therefore, for a fixed γ, there exists a sufficiently small $\bar{\alpha}$ such that $g_j(x^* + \alpha d_\gamma) \leq 0$ for all j and $\alpha \in (0, \bar{\alpha}]$. Thus $d_\gamma \in F(x^*)$ for all γ and

$$\lim_{\gamma \to 0} d_\gamma = d \in \overline{F}(x^*).$$

(3) Since g_j is convex, we have for every $j \in A(x^*)$

$$g_j(x^*) + \nabla g_j(x^*)'(\overline{x} - x^*) \leq g_j(\overline{x}) < 0.$$

By defining $d = \overline{x} - x^*$ and by using $g_j(x^*) = 0$ for all $j \in A(x^*)$, from the preceding relation we obtain

$$\nabla g_j(x^*)d < 0, \qquad \forall\, j \in A(x^*),$$

and the result follows from part (2).

(4) Let B be a matrix with rows consisting of $\nabla g_j(x^*)'$ for $j \in A(x^*)$. Since these gradients are linearly independent, B has full row rank, so that the square matrix BB' is invertible and the matrix $B_r = B'(BB')^{-1}$ is well-defined. Let

$$d = B_r \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix}.$$

Multiplying both sides of this equation with B, we obtain

$$Bd = \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix},$$

which is equivalent to

$$\nabla g_j(x^*)'d = -1, \qquad \forall\, j \in A(x^*).$$

The result now follows from part (2).

(d) For this problem we can easily see that the point $x^* = (0,0)$ is a constrained local minimum. We have

$$\nabla g_1(0,0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{and} \quad \nabla g_2(0,0) = \begin{pmatrix} 0 \\ -1 \end{pmatrix}.$$

Note that both constraints are active at $x^* = (0,0)$, i.e., $A(x^*) = \{1,2\}$. Evidently g_1 and g_2 are not linear, so the condition (c1) does not hold. Furthermore, there is no vector $d = (d_1, d_2)'$ such that

$$\nabla g_1(0,0)'d = d_2 < 0 \quad \text{and} \quad \nabla g_2(0,0)'d = -d_2 < 0.$$

Hence, the condition (c2) is violated. If the condition (c3) holds, then as seen in proof of part (c3) the condition (c2) also holds, which is a contradiction. Therefore, at $x^* = (0,0)$ the condition (c3) does not hold. The vectors $\nabla g_1(0,0)$ and $\nabla g_2(0,0)$ are linearly dependent since $\nabla g_1(0,0) = -\nabla g_2(0,0)$, so the condition (c4) is also violated.

Let scalars $\mu_0 \geq 0$, $\mu_1 \geq 0$, and $\mu_2 \geq 0$ be such that

$$\mu_0 \nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) = 0,$$

or equivalently

$$\begin{pmatrix} \mu_0 \\ \mu_0 \end{pmatrix} + \begin{pmatrix} 0 \\ \mu_1 \end{pmatrix} + \begin{pmatrix} 0 \\ -\mu_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

It follows that $\mu_0 = 0$, i.e., there is no Lagrange multiplier.

(e) Note that $\{x \mid h(x) = 0\} = \{x \mid ||h(x)||^2 \leq 0\}$, so that x^* is also a local minimum for the modified problem. The modified problem has a single constraint $g_1(x) = ||h(x)||^2$, which is active at x^*. Since g_1 is not linear, the condition (c1) does not hold. Because $\nabla g_1(x^*) = 2\nabla h(x^*) h(x^*) = 0$, the conditions (c2) and (c4) are violated at x^*. If g_1 is convex and the condition (c3) holds, then as seen in the proof of (c3), the condition (c2) also holds, which is a contradiction. Hence, at x^* each of the conditions (1)–(4) of part (c) is violated. From

$$\mu_0^* \nabla f(x^*) + \mu_1^* \nabla g_1(x^*) = 0$$

and $\nabla g_1(x^*) = 0$, it follows that $\mu_0^* \nabla f(x^*) = 0$, and since $\nabla f(x^*) \neq 0$, we must have $\mu_0^* = 0$, i.e., there is no Lagrange multiplier.

3.3.6 www

Assume that there exist $x \in \Re^n$ and $\mu \in \Re^m$ such that conditions (i) and (ii) hold, i.e.,

$$a_i' x < 0, \qquad \forall\, i = 1, \ldots, m, \tag{1}$$

$$\sum_{i=1}^{m} \mu_i a_i = 0, \quad \mu \neq 0, \quad \mu \geq 0, \tag{2}$$

where a_i' are row vectors of the matrix A. Without loss of generality, we may assume that $\mu_1 > 0$. By pre-multiplying Eq. (1) with $\mu_i \geq 0$ and summing the obtained inequalities over i, we have

$$\sum_{i=1}^{m} \mu_i a_i' x \leq \mu_1 a_1' x < 0.$$

On other hand, from Eq. (2) we obtain

$$\sum_{i=1}^{m} \mu_i a_i' x = 0,$$

which is a contradiction. Hence, conditions (i) and (ii) cannot hold simultaneously.

The proof will be complete if we can show that conditions (i) and (ii) cannot *fail* to hold simultaneously. Indeed, if condition (i) fails to hold, the minimax problem

$$\begin{aligned}&\text{minimize} \ \ \max\{a_1'x, \ldots, a_m'x\}\\ &\text{subject to} \ \ x \in \Re^n\end{aligned}$$

has 0 as its solution. Hence by Prop. 3.3.10, there exists a $\mu \geq 0$ with $\sum_{i=1}^m \mu_i = 1$ such that $\sum_{i=1}^m \mu_i a_i = 0$, or $A'\mu = 0$. Thus condition (ii) holds, and it follows that the conditions (i) and (ii) cannot fail to hold simultaneously.

3.3.7 (www)

Assume, to obtain a contradiction, that the conclusion does not hold, so that there is a sequence $\{x^k\}$ such that $x^k \to x^*$, and for all k, $x^k \neq x^*$, $h(x^k) = 0$, and $f(x^k) < f(x^*) + (1/k)||x^k - x^*||^2$. Let us write $x^k = x^* + \delta^k y^k$, where

$$\delta^k = \|x^k - x^*\|, \qquad y^k = \frac{x^k - x^*}{\|x^k - x^*\|}.$$

The sequence $\{y^k\}$ is bounded and lies on the surface of the unit sphere, so it must have a subsequence converging to some y with $\|y\| = 1$. Without loss of generality, we assume that the whole sequence $\{y^k\}$ converges to y.

By taking the limit as $\delta^k \to 0$ in the relations

$$\frac{1}{k}||x^k - x^*|| > \frac{f(x^* + \delta^k y^k) - f(x^*)}{\delta^k} = \nabla f(x^*)'y^k + \frac{o(\delta^k)}{\delta^k},$$

$$0 = \frac{h_i(x^k) - h_i(x^*)}{\delta^k} = \frac{h_i(x^* + \delta^k y^k) - h_i(x^*)}{\delta^k} = \nabla h_i(x^*)'y^k + \frac{o(\delta^k)}{\delta^k},$$

$$0 \geq \frac{g_j(x^k) - g_j(x^*)}{\delta^k} = \frac{g_j(x^* + \delta^k y^k) - g_j(x^*)}{\delta^k} = \nabla g_j(x^*)'y^k + \frac{o(\delta^k)}{\delta^k},$$

we see that

$$\nabla f(x^*)'y \leq 0, \qquad \nabla h(x^*)'y = 0, \quad i = 1, \ldots, m, \quad \nabla g_j(x^*)'y \leq 0, \qquad \forall\ j \in A(x^*).$$

Let us now show that

$$\nabla g_j(x^*)'y = 0, \qquad \forall\ j \in A^+(x^*), \tag{1}$$

where

$$A^+(x^*) = \{j \mid \mu_j^* > 0\},$$

so that we can conclude based on the hypothesis that

$$y'\nabla^2_{xx}L(x^*,\lambda^*)y > 0. \tag{2}$$

Indeed, we have $\nabla_x L(x^*,\lambda^*,\mu^*) = 0$ or equivalently

$$\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j\in A^+(x^*)} \mu_j^* \nabla g_j(x^*) = 0.$$

By taking inner product of this relation with y and by using the equation $\nabla h_i(x^*)'y = 0$, we obtain

$$\nabla f(x^*)'y + \sum_{j\in A^+(x^*)} \mu_j^* \nabla g_j(x^*)'y = 0.$$

Since all the terms in the above equation have been shown to be nonpositive, they must all be equal to 0, showing that Eq. (1) holds.

We will now show that $y'\nabla^2_{xx}L(x^*,\lambda^*)y \le 0$, thus coming to a contradiction [cf. Eq. (2)]. Since $x^k = x^* + \delta^k y^k$, by the mean value theorem [Prop. A.23(b) in Appendix A], we have

$$\frac{1}{k}||x^k - x^*||^2 > f(x^k) - f(x^*) = \delta^k \nabla f(x^*)'y^k + \frac{(\delta^k)^2}{2} y^{k\prime}\nabla^2 f(\tilde{\xi}^k) y^k, \tag{3}$$

$$0 = h_i(x^k) - h_i(x^*) = \delta^k \nabla h_i(x^*)'y^k + \frac{(\delta^k)^2}{2} y^{k\prime}\nabla^2 h_i(\overline{\xi}_i^k) y^k, \qquad i = 1,\ldots,m, \tag{4}$$

$$0 \ge g_j(x^k) - g_j(x^*) = \delta^k \nabla g_j(x^*)'y^k + \frac{(\delta^k)^2}{2} y^{k\prime}\nabla^2 g_j(\hat{\xi}_j^k) y^k, \qquad j \in A(x^*), \tag{5}$$

where all the vectors $\tilde{\xi}^k$, $\overline{\xi}_i^k$, and $\hat{\xi}_j^k$ lie on the line segment joining x^* and x^k. Multiplying Eqs. (4) and (5) by λ_i^* and μ_j^*, respectively, adding them and adding Eq. (3) to them, we obtain

$$\begin{aligned}\frac{1}{k}||x^k - x^*||^2 > \delta^k &\left(\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j\in A(x^*)} \mu_j^* \nabla g_j(x^*)\right)' y^k \\ &+ \frac{(\delta^k)^2}{2} y^{k\prime} \left(\nabla^2 f(\tilde{\xi}^k) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(\overline{\xi}_i^k) + \sum_{j\in A(x^*)} \mu_j^* \nabla^2 g_j(\hat{\xi}_j^k)\right) y^k.\end{aligned}$$

Since $\delta^k = ||x^k - x^*||$ and $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j\in A(x^*)} \mu_j^* \nabla g_j(x^*) = 0$, we obtain

$$\frac{2}{k} > y^{k\prime} \left(\nabla^2 f(\tilde{\xi}^k) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(\overline{\xi}_i^k) + \sum_{j\in A(x^*)} \mu_j^* \nabla^2 g_j(\hat{\xi}_j^k)\right) y^k.$$

By taking the limit as $k \to \infty$,

$$0 \ge y' \left(\nabla^2 f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla^2 h_i(x^*) + \sum_{j\in A(x^*)} \mu_j^* \nabla^2 g_j(x^*)\right) y,$$

thus arriving at the desired contradiction.

3.3.10 (www)

(a) Consider a problem where there are two identical equality constraints [$h_1(x) = h_2(x)$ for all x], and assume that x^* is a local minimum such that $\nabla h_1(x^*) \neq 0$. Then, $\nabla f(x^*) + \lambda \nabla h_1(x^*) = 0$ for some λ. Take a scalar $\gamma > 0$ such that $\lambda + \gamma > 0$ and let $\lambda_1^* = \lambda + \gamma$ and $\lambda_2^* = -\gamma$. Then we have

$$\nabla f(x^*) + \lambda_1^* \nabla h_1(x^*) + \lambda_2^* \nabla h_2(x^*) = 0,$$

but since λ_1^* and λ_2^* have different signs, there is no x such that simultaneously we have $\lambda_1^* h_1(x) > 0$ and $\lambda_2^* h_2(x) > 0$. Thus λ_1^* and λ_2^* violate the last Fritz John condition. As an alternative example, consider the following inequality constrained problem

$$\begin{aligned} &\text{minimize} \quad x_1 + x_2 \\ &\text{subject to} \;\; g_1(x_1, x_2) = (x_1)^2 - x_2 \leq 0, \quad g_2(x_1, x_2) = -(x_1)^2 + x_2 \leq 0. \end{aligned}$$

Then $x^* = (0, 0)$ is a local minimum with $A(x^*) = \{1, 2\}$, and $\mu_0^* = \mu_1^* = \mu_2^* = 1$ satisfy Karush-Kun-Tucker conditions, namely

$$\nabla f(0, 0) + \nabla g_1(0, 0) + \nabla g_2(0, 0) = 0.$$

However, there is no point (x_1, x_2) such that $g_1(x_1, x_2) > 0$ and $g_2(x_1, x_2) > 0$, i.e., the Fritz John condition (iv) does not hold.

(b) For simplicity, assume that all the constraints are inequalities (equality constraints can be handled by conversion to two inequalities). If $\nabla f(x^*) = 0$, we can take $\mu_j = 0$ for all j, and we are done. Assume that $\nabla f(x^*) \neq 0$ and consider the index subsets $J \subset A(x^*)$ such that $\nabla f(x^*)$ is a positive combination of the gradients $\nabla g_j(x^*)$, $j \in J$, and among all such subsets, let J^* have a minimal number of elements. Without loss of generality, let $J^* = \{1, \ldots, s\}$, so we have

$$\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \cdots + \mu_s \nabla g_s(x^*) = 0,$$

where $\mu_j > 0$ for $j = 1, \ldots, s$.

We claim that $\nabla g_1(x^*), \ldots, \nabla g_s(x^*)$ are linearly independent. Indeed, if this were not so, we would have for some $\alpha_1, \ldots, \alpha_s$, not all zero,

$$\alpha_1 \nabla g_1(x^*) + \cdots + \alpha_s \nabla g_s(x^*) = 0$$

so that

$$\nabla f(x^*) + (\mu_1 + \gamma \alpha_1) \nabla g_1(x^*) + \cdots + (\mu_s + \gamma \alpha_s) \nabla g_s(x^*) = 0,$$

for all scalars γ. Thus, we can find γ such that $\mu_j + \gamma\alpha_j \geq 0$ for all j and $\mu_{\overline{j}} + \gamma\alpha_{\overline{j}} = 0$ for at least one index $\overline{j} \in \{1, \ldots, r\}$. This contradicts the hypothesis that the index set J^* has a minimal number of elements.

Thus $\nabla g_1(x^*), \ldots, \nabla g_s(x^*)$ are linearly independent, so that we can find a vector h such that

$$\nabla g_1(x^*)'h = \cdots = \nabla g_s(x^*)'h = 1.$$

Consider vectors of the form

$$x = x^* + \gamma h,$$

where γ is a positive scalar. By Taylor's theorem, for sufficiently small γ, we have $g_j(x^* + \gamma h) > 0$ and hence also $\mu_j g_j(x^* + \gamma h) > 0$ for all $j = 1, \ldots, s$. Thus, the scalars μ_j, $j = 1, \ldots, s$, together with $\mu_j = 0$ for $j = s + 1, \ldots, r$, satisfy all the Fritz John conditions with $\mu_0 = 1$.

3.3.11 (www)

From the given conditions, it follows that

$$\sum_{j \in A(x^*)} \mu_j^* \nabla g_j(x^*) = 0, \tag{1}$$

where $\mu_1^*, \ldots, \mu_r^*$ are Lagrange multipliers satisfying the Fritz John conditions. Since the functions $g_j(x)$ are convex over $\Re^n$, for any $j \in A(x^*)$ and any feasible vector x we have

$$0 \geq g_j(x) - g_j(x^*) \geq \nabla g_j(x^*)'(x - x^*).$$

Therefore

$$\begin{aligned} \mu_j^* g_j(x) &\geq \mu_j^* \big(g_j(x^*) + \nabla g_j(x^*)'(x - x^*) \big) \\ &= \mu_j^* \nabla g_j(x^*)'(x - x^*), \quad \forall\ j \in A(x^*). \end{aligned}$$

This and Eq. (1) imply

$$\sum_{j \in A(x^*)} \mu_j^* g_j(x) \geq 0, \quad \text{for all feasible } x.$$

On the other hand, for all feasible x we have $\sum_{j \in A(x^*)} \mu_j^* g_j(x) \leq 0$. Therefore

$$\sum_{j \in A(x^*),\ \mu_j^* > 0} \mu_j^* g_j(x) = \sum_{j \in A(x^*)} \mu_j^* g_j(x) = 0$$

for all feasible x. This is possible only if $g_j(x) = 0$ for all feasible x and $j \in A(x^*)$ with $\mu_j^* > 0$. Since not all μ_j^* are equal to zero, there is at least one index j with $\mu_j^* > 0$.

3.3.12 (www)

It is straightforward that the given condition is implied by the condition (iv) of Prop. 3.3.5.

To show the reverse, we replace each equality constraint $h_i(x) = 0$ with the two constraints $h_i(x) \leq 0$ and $-h_i(x) \leq 0$, and we apply the version of the Fritz John conditions given in the exercise. Let λ_i^+ and λ_i^- be the multipliers corresponding to the constraints $h_i(x) \leq 0$ and $-h_i(x) \leq 0$, respectively. Thus in any neighborhood N of x^* there is a vector x such that

$$h_i(x) > 0, \qquad \text{for all } \ i \ \text{ with } \lambda_i^+ > 0, \tag{1}$$

$$-h_i(x) > 0, \qquad \text{for all } \ i \ \text{ with } \lambda_i^- > 0, \tag{2}$$

$$g_j(x) > 0, \qquad \text{for all } \ j \ \text{ with } \mu_j^* > 0.$$

Evidently $\mu_j^* g_j(x) > 0$ for all j with $\mu_j^* > 0$. Since $\lambda_i^* = \lambda_i^+ - \lambda_i^-$, if $\lambda_i^* \neq 0$ then either $\lambda_i^+ > \lambda_i^- = 0$ (corresponds to $\lambda_i^* > 0$) or $\lambda_i^- > \lambda_i^+ = 0$ (corresponds to $\lambda_i^* < 0$). In either case, from Eqs. (1) and (2) we have that

$$\lambda_i^* h_i(x) > 0, \qquad \text{for all } i \ \text{ with } \ \lambda_i^* \neq 0.$$

Hence the Fritz John condition (iv), as given in Prop. 3.3.5, holds.

3.3.13 (www)

First, let us point out some important properties of a convex function that will be used in the proof.

Convexity of f over $\Re^n$ implies that f is continuous over $\Re^n$ and the set $\partial f(x)$ of subgradients of f at x is nonempty for all $x \in \Re^n$ (see Prop. B.24 of Appendix B).

If f is convex over $\Re^n$, while G is continuously differentiable over $\Re^n$, then if a point y^* is an unconstrained local minimum of $f(x) + G(x)$, we have if $0 \in \partial f(y^*) + \nabla G(y^*)$ (see Prop. B.24 of Appendix B).

(a) Let x^* be a local minimum of f and $S = \{x \mid ||x - x^*|| \leq \epsilon\}$, where $\epsilon > 0$ is such that $f(x) \geq f(x^*)$ for all feasible x with $x \in S$. As in the proof of Prop. 3.1.1 (Sec. 3.1.1), for each $k \geq 1$ we consider the penalized problem

$$\begin{aligned} &\text{minimize} \ \ F^k(x) = f(x) + \frac{k}{2}\sum_{i=1}^{m}(h_i(x))^2 + \frac{k}{2}\sum_{j=1}^{r}(g_j^+(x))^2 + \frac{1}{2}||x - x^*||^2 \\ &\text{subject to} \ \ x \in S. \end{aligned}$$

Similar to Sec. 3.1.1, we conclude that the solution x^k for the above problem exists and (using the continuity of f, h_i, g_j^+) that $x^k \to x^*$ as $k \to \infty$. Therefore, there is an index $\overline{k}$ such that x^k is an interior point of S for all $k \geq \overline{k}$. For such k, we have $0 \in \partial F^k(x^k)$, or equivalently

$$s^k + \sum_{i=1}^m \xi_i^k \nabla h_i(x^k) + \sum_{j=1}^r \zeta_j^k \nabla g_j(x^k) + (x^k - x^*) = 0,$$

for some $s^k \in \partial f(x^k)$ and $\xi_i^k = k h_i(x^k)$, $\zeta_j^k = k g_j^+(x^k)$.

Following the lines of the proof of Prop. 3.3.5, we obtain

$$\mu_0^k s^k + \sum_{i=1}^m \lambda_i^k \nabla h_i(x^k) + \sum_{j=1}^r \mu_j^k \nabla g_j(x^k) + \frac{1}{\delta^k}(x^k - x^*) = 0,$$

for all $k \geq \overline{k}$, where

$$\mu_0^k = \frac{1}{\delta^k}, \qquad \lambda_i^k = \frac{\xi_i^k}{\delta^k}, \quad i = 1, \ldots, m, \qquad \mu_j^k = \frac{\zeta_j^k}{\delta^k}, \quad j = 1, \ldots, r,$$

and

$$\delta^k = \sqrt{1 + \sum_{i=1}^m (\xi_i^k)^2 + \sum_{j=1}^r (\zeta_j^k)^2}.$$

Since $x^k \to x^*$ with $s^k \in \partial f(x^k)$ for all k, from Prop. B.24 and the boundedness of the sequence $\{\mu_0^k, \lambda_1^k, \ldots, \lambda_m^k, \mu_1^k, \ldots, \mu_r^k\}$ we see that there are a vector $s^* \in \partial f(x^*)$ and a limit point $(\mu_0^*, \lambda_1^*, \ldots, \lambda_m^*, \mu_1^*, \ldots, \mu_r^*)$ such that

$$\mu_0^* s^* + \sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*) = 0, \tag{1}$$

If $\mu^* = 0$, then the vector

$$-\sum_{i=1}^m \lambda_i^* \nabla h_i(x^*) - \sum_{j=1}^r \mu_j^* \nabla g_j(x^*)$$

is equal to zero. Otherwise, we can set $\mu_0^* = 1$ in (1), which shows that the above vector is a subgradient of f at x^*. Thus, condition (i) of the exercise is satisfied. The rest of the proof is the same as that of Prop. 3.3.5.

(b) The proof is similar to the one of Prop. 3.3.7.

(b) Assume that $\nabla h_i(x^*)$ are linearly independent, and that there is a vector d such that

$$\nabla h_i(x^*)'d = 0, \qquad \forall\ i = 1, \ldots, m, \qquad \nabla g_j(x^*)'d < 0, \qquad \forall\ j \in A(x^*).$$

If $\mu_0^* = 0$ in (1), then using the same argument as in proof of Prop. 3.3.8 we arrive at contradiction. Under the Slater condition, the proof that $\mu_0^* \neq 0$ is the same as in Prop. 3.3.9.

3.3.14 (www)

The problem can be formulated as follows

$$\text{minimize } r^2$$
$$\text{subject to } ||x - y_j||^2 \leq r^2, \quad j = 1, \ldots, p, \quad x \in \Re^n,$$

which is equivalent to the unconstrained minimax problem

$$\text{minimize } \max\{||x - y_1||^2, \ldots, ||x - y_p||^2\}$$
$$\text{subject to } x \in \Re^n.$$

According to Prop. 3.3.10, the Lagrange multiplier conditions are

(i) $2\sum_{j=1}^{p} \mu_j^*(x^* - y_j) = 0.$

(ii) $\mu^* \geq 0, \qquad \sum_{j=1}^{p} \mu_j^* = 1.$

(iii) For all $j = 1, \ldots, p$, if $\mu_j^* > 0$, then

$$||x^* - y_j||^2 = \max\{||x - y_1||^2, \ldots, ||x - y_p||^2\},$$

where x^* is optimal solution for the minimax problem and μ^* is the corresponding Lagrange multiplier.

Note that the cost function is continuous and coercive, so that the optimal solution always exists. Furthermore, the cost function is convex and the given conditions are also sufficient for optimality. By combining (i) and (ii) we have

$$x^* = \sum_{j=1}^{p} \mu_j^* y_j, \quad \sum_{j=1}^{p} \mu_j^* = 1, \quad \mu_j^* > 0, \ \ \forall\, j,$$

i.e., x^* is a convex combination of the given points $y_1, \ldots, y_p$. For $p = 3$, when y_1, y_2, y_3 do not lie on the same line, we have the following geometric solution:

(1) All constraints are active, so x^* is at equal distance from all three points. Then x^* is the center of the circle circumscribed around the triangle of the three points. In this case x^* must lie within the triangle and is a positive combination of the y_j, the coefficients being the multipliers. This corresponds to the case when the triangle is not obtuse.

(2) Only two of the constraints are active, in which case x^* lies on the line connecting the two points. This occurs when the triangle formed by the given points is obtuse. Then x^* is the midpoint of the longest side of the triangle. If y_j is not the end point of the longest side, then $\mu_j = 0$. The other two Lagrange multipliers are both positive.

Now consider the degenerate case when the three points lie on the same line. We can assume that y_3 lies between y_1 and y_2. Then the optimal point x^* is the midpoint of the segment joining y_1 and y_2. The Lagrange multipliers μ_1^* and μ_2^* are positive, while $\mu_3^* = 0$.

3.3.15 (www)

(a) Let $\{y^k\}$ be a sequence of points in $T(x)$ for some $x \in X$. Assume that $y^k \to y$ as $k \to \infty$. The definition of the tangent cone implies that for every y^k there is a sequence $\{x_i^k\} \subseteq X \setminus \{x\}$ such that

$$x_i^k \to x, \quad \text{and} \quad \frac{x_i^k - x}{||x_i^k - x||} \to \frac{y^k}{||y^k||} \quad \text{as} \quad i \to \infty.$$

For $k = 1, 2, \ldots$, choose an index i_k such that $i_k > i_{k-1} > \ldots > i_1$ and

$$||x_{i_k}^k - x|| < \frac{1}{2^k} \quad \text{and} \quad \left\| \frac{x_{i_k}^k - x}{||x_{i_k}^k - x||} - \frac{y^k}{||y^k||} \right\| < \frac{1}{2^k}.$$

Evidently $\{x_{i_k}^k\} \subseteq X \setminus \{x\}$, and $x_{i_k}^k \to x$ as $k \to \infty$. Also, we have that $\left\| \frac{x_{i_k}^k - x}{||x_{i_k}^k - x||} - \frac{y^k}{||y^k||} \right\| \to 0$ as $k \to \infty$. This together with the fact that $y^k \to y$, and

$$\left\| \frac{x_{i_k}^k - x}{||x_{i_k}^k - x||} - \frac{y}{||y||} \right\| \le \left\| \frac{x_{i_k}^k - x}{||x_{i_k}^k - x||} - \frac{y^k}{||y^k||} \right\| + \left\| \frac{y^k}{||y^k||} - \frac{y}{||y||} \right\|,$$

implies

$$\lim_{k \to \infty} \left\| \frac{x_{i_k}^k - x}{||x_{i_k}^k - x||} - \frac{y}{||y||} \right\| = 0,$$

which by the definition of $T(x)$ means that $y \in T(x)$. Thus, $T(x)$ is closed.

(b) Let $F(x)$ and $\overline{F}(x)$ denote, respectively, the set of feasible directions at x and its closure. First, we will prove that $\overline{F}(x) \subseteq T(x)$ holds, regardless of whether X is convex. Let $d \in F(x)$. Then there is an $\overline{\alpha} > 0$ such that $x + \alpha d \in X$ for all $\alpha \in [0, \overline{\alpha}]$. Choose any sequence $\{\alpha^k\} \subseteq (0, \overline{\alpha}]$ with $\alpha^k \to 0$ as $k \to \infty$. Define $x^k = x + \alpha^k d$. Evidently $x^k \in X \setminus \{x\}$, and $\frac{x^k - x}{||x^k - x||} = \frac{d}{||d||}$ converges to $\frac{d}{||d||}$. Hence $d \in T(x)$. It follows that $F(x) \subseteq T(x)$, and since $T(x)$ is closed, we have $\overline{F}(x) \subseteq T(x)$.

Next, we prove that $T(x) \subseteq \overline{F}(x)$. Let $y \in T(x)$ and $\{x^k\} \subseteq X \setminus \{x\}$ be such that

$$\frac{x^k - x}{||x^k - x||} = \frac{y}{||y||} + \xi^k,$$

where $\xi^k \to 0$ as $k \to \infty$. Since X is a convex set, the direction $x^k - x$ is feasible at x for all k. Therefore, the direction $d^k = \frac{x^k - x}{||x^k - x||} \cdot ||y|| = y + \xi^k ||y||$ is feasible at x for all k, i.e., $\{d^k\} \subseteq F(x)$. Since

$$\lim_{k \to \infty} d^k = \lim_{k \to \infty} (y + \xi^k ||y||) = y,$$

we have $y \in \overline{F}(x)$. Consequently $T(x) \subseteq \overline{F}(x)$. This completes the proof.

3.3.16 www

Let x be any vector in X. We will show that $T(x) = V(x)$. We have, in general $T(x) \subset V(x)$ (see e.g., the proof of Prop. 3.3.17), so we focus on showing that $V(x) \subset T(x)$. Let $y \in V(x)$, so that we have

$$\nabla g_j(x)'y \le 0, \qquad \forall\ j \in A(x).$$

Let α^k be a positive sequence with $\alpha^k \to 0$, and let

$$x^k = x + \alpha^k y.$$

For all $j \in A(x)$ we have $g_j(x) = 0$, and using the concavity of g_j, we obtain

$$g_j(x^k) \le g_j(x) + \alpha^k \nabla g_j(x)'y \le 0.$$

It follows that for k sufficiently large, x^k is feasible. Since

$$x^k \to x, \qquad \frac{x^k - x}{\|x^k - x\|} = \frac{y}{\|y\|},$$

it follows that $y \in T(x)$, so that $V(x) \subset T(x)$.

3.3.17 www

Let y be a vector such that $\nabla g_j(x^*)'y < 0$ for all $j \in A(x^*)$. By continuity of $\nabla g_j(x)$ (as a function of x and j), there exist a neighborhood N of x^* and a neighborhood A of $A(x^*)$ (relative to J) such that

$$\nabla g_j(x)'y < 0, \quad \forall\ x \in N, \quad \forall\ j \in A. \tag{1}$$

Furthermore, the neighborhood N can be chosen so that

$$g_j(x) < 0, \quad \forall\ x \in N, \quad \forall\ j \in J \setminus A. \tag{2}$$

Since N is open and $x^* \in N$, we can find a scalar $\overline{\alpha} > 0$ so that $x^* + \alpha y \in N$ whenever $0 \le \alpha \le \overline{a}$. For any α with $0 < \alpha \le \overline{a}$ and $j \in A$, by the mean value theorem and feasibility of x^*, we have

$$g_j(x^* + \alpha y) = g_j(x^*) + \alpha \nabla g_j(x^* + \theta\alpha y)'y \le \alpha \nabla g_j(x^* + \theta\alpha y)'y, \tag{3}$$

for some $\theta \in (0,1)$. Since $x^* + \theta\alpha y \in N$ and $j \in A$, from Eqs. (1) and (3) we obtain

$$g_j(x^* + \alpha y) < 0, \quad \forall\ j \in A, \quad \forall\ \alpha \in (0,1].$$

For any α with $0 < \alpha \le \overline{a}$ the point $x^* + \alpha y$ belongs to N, which together with Eq. (2) implies

$$g_j(x^* + \alpha y) < 0, \quad \forall\ j \in J \setminus A, \quad \forall\ \alpha \in (0,1].$$

The last two inequalities show that y is a feasible direction of X at x^*. In the solution to part (b) of Exercise 3.3.15, it is shown that the set of feasible directions at x^* is a subset of the tangent cone at x^*, regardless of the structure of the set X.

3.3.18 (www)

Assume that we have shown the validity of the Mangasarian-Fromovitz constraint qualification for the problem without equality constraints, i.e., for a local minimum x^*, there exist Lagrange multipliers under the condition that there is a vector d such that

$$\nabla g_j(x^*)'d < 0, \qquad \forall\ j \in A(x^*). \tag{1}$$

Now, consider the problem with equality and inequality constraints. Assume that there is a vector d such that

$$\begin{aligned} &\nabla h_i(x^*)'d = 0, \qquad \forall\ i = 1, \ldots, m, \\ &\nabla g_j(x^*)'d < 0, \qquad \forall\ j \in A(x^*). \end{aligned} \tag{2}$$

Since the vectors $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ are linearly independent, by reordering the coordinates of x if necessary, we can partition the vector x as $x = (x_B, x_R)$ such that the submatrix $\nabla_B h(x^*)$ (the gradient matrix of h with respect to x_B) is invertible. The equation

$$h(x_B, x_R) = 0$$

has the solution (x_B^*, x_R^*), and the implicit function theorem (Prop. A.25 of Appendix A) can be used to express x_B in terms of x_R via a unique continuously differentiable function $\phi : S \mapsto \Re^m$ defined over a sphere S centered at x_R^*. In particular, we have $x_B^* = \phi(x_R^*)$, $h\left(\phi(x_R), x_R\right) = 0$ for all $x_R \in S$, and

$$\nabla\phi(x_R) = -\nabla_R h\left(\phi(x_R), x_R\right) \left(\nabla_B h\left(\phi(x_R), x_R\right)\right)^{-1}, \quad \forall\ x_R \in S, \tag{3}$$

where $\nabla_R h$ is the gradient matrix of h with respect to x_R. Observe that x_R^* is a local minimum of the problem

$$\begin{aligned} &\min F(x_R) \\ &\text{subject to } G_j(x_R) \le 0, \quad j = 1, \ldots, r, \end{aligned} \tag{4}$$

where $F(x_R) = f\left(\phi(x_R), x_R\right)$, $G_j(x_R) = g_j\left(\phi(x_R), x_R\right)$. Note that this problem has no equality constraints. From (2) we have

$$\nabla h(x^*)'d = \nabla_B h(x^*)'d_B + \nabla_R h(x^*)'d_R = 0,$$

and

$$\nabla g_j(x^*)'d = \nabla_B g_j(x^*)'d_B + \nabla_R g_j(x^*)'d_R < 0, \tag{5}$$

for all $j \in A(x^*)$. Since $\nabla_B h(x^*)'$ is invertible, from the first relation above we obtain

$$d_B = -\left(\nabla_B h\left(\phi(x_R^*), x_R^*\right)'\right)^{-1} \nabla_R h\left(\phi(x_R^*), x_R^*\right)' d_R,$$

which in view of Eq. (3) is equivalent to

$$d_B = \nabla\phi(x_R^*)' d_R.$$

Substituting this in Eq. (5), we obtain

$$\nabla_B g_j\left(\phi(x_R^*), x_R^*\right)' \nabla\phi(x_R^*)' d_R + \nabla_R g_j\left(\phi(x_R^*), x_R^*\right)' d_R < 0,$$

which is equivalent to

$$\nabla G_j(x_R^*)' d < 0, \qquad \forall\ j \in A(x^*).$$

This means that the Mangasarian-Fromovitz constraint qualification is satisfied for problem (4), so there are Lagrange multipliers $\mu_1^*, \ldots, \mu_r^*$ such that

$$\begin{aligned}
0 = \nabla F(x_R^*) + \sum_{j=1}^r \mu_j^* \nabla G_j(x_R^*) &= \nabla\phi(x_R^*)\nabla_B f(x^*) + \nabla_R f(x^*) \\
&\quad + \sum_{j=1}^r \mu_j^* \left(\nabla\phi(x_R^*)\nabla_B g_j(x^*) + \nabla_R g_j(x^*)\right) \\
&= \nabla\phi(x_R^*) \left(\nabla_B f(x^*) + \sum_{j=1}^r \mu_j^* \nabla_B g_j(x^*)\right) + \nabla_R f(x^*) \\
&\quad + \sum_{j=1}^r \mu_j^* \nabla_R g_j(x^*).
\end{aligned} \tag{6}$$

Define

$$B' = \nabla_B h\left(\phi(x_R^*), x_R^*\right), \quad R' = \nabla_R h\left(\phi(x_R^*), x_R^*\right)$$

and

$$\lambda^* = -B'^{-1}\left(\nabla_B f(x^*) + \sum_{j=1}^r \mu_j^* \nabla_B g_j(x^*)\right).$$

Then from Eq. (3) we see that $\nabla\phi(x_R^*) = -R'B'^{-1}$, which combined with Eq. (6) implies

$$\nabla_R f(x^*) + R'\lambda^* + \sum_{j=1}^r \mu_j^* \nabla_R g_j(x^*) = 0.$$

The definition of λ^* implies

$$\nabla_B f(x^*) + B'\lambda^* + \sum_{j=1}^r \mu_j^* \nabla_B g_j(x^*) = 0.$$

Since $\nabla h(x^*)' = (B', R')$, the last two equalities are equivalent to

$$\nabla f(x^*) + \nabla h(x^*)'\lambda^* + \sum_{j=1}^r \mu_j^* \nabla g_j(x^*) = 0,$$

which shows that the Lagrange multipliers exist.

The proof of the existence of the Lagrange multipliers under the Slater constraint qualification is straightforward from the preceding analysis by noting that the vector $d = \overline{x} - x^*$ satisfies the Mangasarian-Fromovitz constraint qualification.

3.3.19 (www)

For simplicity we assume that there are no equality constraints; the subsequent proof can be easily extended to the case whether there are some inequality constraints. To show that the Mangasarian-Fromovitz constraint qualification implies boundedness of the set of Lagrange multipliers, follow the given hint.

Conversely, if the set of Lagrange multipliers is bounded, there cannot exist a $\mu \neq 0$ with $\mu \geq 0$ and $\sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0$, since adding $\gamma\mu$, for any $\gamma > 0$, to a Lagrange multiplier gives another Lagrange multiplier. Hence by the theorem of the alternative of Exercise 3.3.6, there must exist a d such that $\nabla g_j(x^*)'d < 0$ for all $j \in A(x^*)$.

3.3.20 (www)

We have

$$\nabla h_1(x) = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

$$\nabla h_2(x) = \begin{cases} \begin{pmatrix} 4x_1^3 \sin\left(\frac{1}{x_1}\right) - x_1^2 \cos\left(\frac{1}{x_1}\right) \\ -1 \end{pmatrix} & \text{if } x_1 \neq 0, \\ \begin{pmatrix} 0 \\ -1 \end{pmatrix} & \text{if } x_1 = 0, \end{cases}$$

and it can be seen that ∇h_1 and ∇h_2 are everywhere continuous. Thus, for $\lambda_1 = 1$, $\lambda_2 = 1$, we have

$$\lambda_1 \nabla h_1(0) + \lambda_2 \nabla h_2(0) = 0.$$

On the other hand, it can be seen that arbitrarily closely to $x^* = (0, 0)$, there exists an x such that $h_1(x) > 0$ and $h_2(x) > 0$. Thus x^* is not quasinormal, although it is seen (most easily, by a graphical argument) that x^* is quasiregular.

3.3.21 (www)

(a) Without loss of generality, we assume that there are no equality constraints and that all inequality constraints are active at x^*. Based on the definition of quasinormality, it is easy to verify that x^* is a quasinormal vector of $\overline{X}$ if it is a quasinormal vector of X. Conversely, suppose that x^* is a quasinormal vector of $\overline{X}$, but not a quasinormal vector of X. Then there exist Lagrange multipliers $\mu_1, \ldots, \mu_r$ that satisfy the Fritz John conditions with $\mu_0 = 0$ and $\mu_{\overline{j}} > 0$ for some $\overline{j} \notin \overline{J}$ (for otherwise, x^* would not be a quasinormal vector of $\overline{X}$). From the definition of the set $\overline{J}$ it follows that there is a vector $y \in V(x^*)$ such that $\nabla g_{\overline{j}}(x^*)'y < 0$. By multiplying the relation

$$\sum_{j=1}^{r} \mu_j \nabla g_j(x^*) = 0$$

with y, we obtain

$$0 = \sum_{j=1}^{r} \mu_j \nabla g_j(x^*)'y \le \mu_{\overline{j}} \nabla g_{\overline{j}}(x^*)'y < 0,$$

which is a contradiction. Hence, x^* is a quasinormal vector in X.

(b) Clearly, if x^* is a quasiregular vector of X, then it is a quasiregular vector of $\overline{X}$. To prove the converse, we follow the given hint. Assume that x^* is a quasiregular vector of $\overline{X}$. Then evidently $\tilde{V}(x^*) \subset \overline{V}(x^*) = \overline{T}(x^*)$, where $\overline{V}(x^*)$ and $\overline{T}(x^*)$ denote, respectively, the cone of first order feasible variations and the tangent cone of $\overline{X}$ at x^*. To complete the proof, we need to show that $\tilde{V}(x^*) \subset \overline{T}(x^*)$. Let $y \in \tilde{V}(x^*) \setminus \{0\}$ be arbitrary. Since $y \in \overline{T}(x^*)$, there is a sequence $\{x^k\} \subset \overline{X}$ such that $x^k \neq x^*$ for all k and

$$x^k \to x^*, \qquad \frac{x^k - x^*}{||x^k - x^*||} \to \frac{y}{||y||}.$$

From the first order Taylor's expansion we have

$$\lim_{k \to \infty} \frac{g_j(x^k) - g_j(x^*)}{||x^k - x^*||} = \lim_{k \to \infty} \frac{\nabla g_j(x^*)'(x^k - x^*)}{||x^k - x^*||} = \frac{\nabla g_j(x^*)'y}{||y||}$$

for all j. This implies $g_j(x^k) < 0$ for all $j \notin \overline{J}$ and all sufficiently large k. Therefore $x^k \in X$ for all k sufficiently large, and consequently y is in the tangent cone of X at x^*. Hence $\tilde{V}(x^*) \subset \overline{T}(x^*)$, which is equivalent to quasiregularity of x^* with respect to the set X.

(c) The given statement follows from parts (a) and (b).

3.3.22 (www)

Without loss of generality, we can assume that there are no equality constraints (every equality constraint $h_i(x) = 0$ can be replaced by two inequalities $h_i(x^*) \le 0$ and $-h_i(x^*) \le 0$ with $h_i(x)$ and $-h_i(x)$ being linear, and therefore concave). Since x^* is a local minimum, there exist a scalar μ_0 and Lagrange multipliers $\lambda_1, \ldots, \lambda_m, \mu_1, \ldots, \mu_r$ satisfying the Fritz John conditions. Assume that $\mu_0 = 0$. Then

$$\sum_{j=1}^{r} \mu_j \nabla g_j(x^*) = \sum_{j \in A(x^*)} \mu_j \nabla g_j(x^*) = 0. \tag{1}$$

Multiplying this equation by d, we obtain

$$\sum_{j\in A(x^*)} \mu_j \nabla g_j(x^*)'d = 0. \tag{2}$$

If $\mu_{j_0} > 0$ for some $j_0 \in A(x^*) \setminus J$, then

$$\sum_{j\in A(x^*)} \mu_j \nabla g_j(x^*)'d \le \mu_{j_0} \nabla g_{j_0}(x^*)'d < 0,$$

which is a contradiction to Eq. (2). Therefore for all $j_0 \in A(x^*) \setminus J$ we must have $\mu_j = 0$. Then from Eq. (1) we have

$$\sum_{j\in J} \mu_j \nabla g_j(x^*) = 0. \tag{3}$$

Now we use the same line of argument as in the proof of Prop. 3.3.6 in order to arrive at a contradiction. In particular, since g_j is concave for every $j \in J$, we have

$$g_j(x) \le g_j(x^*) + \nabla g_j(x^*)'(x - x^*), \quad \forall\ j \in J.$$

By multiplying this inequality with μ_j and adding over $j \in J$, we obtain

$$\sum_{j\in J} \mu_j g_j(x) \le \sum_{j\in J} \mu_j g_j(x^*) + \left(\sum_{j\in J} \mu_j \nabla g_j(x^*)\right)' (x - x^*) = 0, \tag{4}$$

where the last equality follows from Eq. (3) and the fact that $\mu_j g_j(x^*) = 0$ for all j [by the Fritz John condition (iv)]. On the other hand, we know that there is some $j \in J$ for which $\mu_j > 0$ and an x satisfying $g_j(x) > 0$ for all j with $\mu_j > 0$. For this x, we have $\sum_{j\in J} \mu_j g_j(x) > 0$, which contradicts Eq. (4). Thus, we can take $\mu_0 = 1$ so that x^* satisfies the necessary conditions of Prop. 3.3.7.

SECTION 3.4

3.4.3 www

Let's first consider

$$(P) \qquad \min_{A'x\ge b} c'x \Longleftrightarrow \max_{A\mu=c,\mu\ge 0} b'\mu. \quad (D)$$

The dual problem to (P) is

$$\max_{\mu \geq 0} q(\mu) = \max_{\mu \geq 0} \inf_{x \in \Re^n} \left\{ \sum_{j=1}^{n} \left(c_j - \sum_{i=1}^{m} \mu_i a_{ij} \right) x_j + \sum_{i=1}^{m} \mu_i b_i \right\}.$$

If $c_j - \sum_{i=1}^{m} \mu_i a_{ij} \neq 0$, then $q(\mu) = -\infty$. Thus the dual problem is

$$\max \sum_{i=1}^{m} \mu_i b_i$$

$$\sum_{i=1}^{m} \mu_i a_{ij} = c_j, \quad j = 1, \ldots, n$$

$$\mu \geq 0.$$

To find the dual of (D), note that (D) is equivalent to

$$\min_{A\mu = c, \mu \geq 0} -b'\mu,$$

and so the dual problem is

$$\max_{x \in \Re^n} p(x) = \max_{x} \inf_{\mu \geq 0} \{(Ax - b)'\mu - c'x\}.$$

If $a_i'x - b_i < 0$ for any i, then $p(x) = -\infty$. Thus the dual of (D) is

$$\max -c'x \quad \text{or} \quad \min c'x$$

$$\text{subject to } A'x \geq b.$$

The Lagrangian optimality condition for (P) is

$$x^* = \arg\min_{x} \left\{ \left(c - \sum_{i=1}^{m} \mu_i^* a_i \right)' x + \sum_{i=1}^{m} \mu_i^* b_i \right\},$$

from which we determine the complementary slackness conditions for (P):

$$A\mu = c.$$

The Lagrangian optimality condition for (D) is

$$\mu^* = \arg\min_{\mu \geq 0} \{(Ax^* - b)'\mu - c'x^*\},$$

from which we determine the complementary slackness conditions for (D):

$$Ax^* - b \geq 0,$$

$$(Ax^* - b)_i \mu_i^* = 0, \quad \forall\ i.$$

Next, consider

$$(P) \qquad \min_{A'x \geq b, x \geq 0} c'x \Longleftrightarrow \max_{A\mu \leq c, \mu \geq 0} b'\mu. \quad (D)$$

The dual problem to (P) is

$$\max_{\mu \geq 0} q(\mu) = \max_{\mu \geq 0} \inf_{x \geq 0} \left\{ \sum_{j=1}^{n} \left(c_j - \sum_{i=1}^{m} \mu_i a_{ij} \right) x_j + \sum_{i=1}^{m} \mu_i b_i \right\}.$$

If $c_j - \sum_{i=1}^m \mu_i a_{ij} < 0$, then $q(\mu) = -\infty$. Thus the dual problem is

$$\max \sum_{i=1}^{m} \mu_i b_i$$

$$\sum_{i=1}^{m} \mu_i a_{ij} \leq c_j, \quad j = 1, \ldots, n$$

$$\mu \geq 0.$$

To find the dual of (D), note that (D) is equivalent to

$$\min_{A\mu \leq c, \mu \geq 0} -b'\mu,$$

and so the dual problem is

$$\max_{x \geq 0} p(x) = \max_{x \geq 0} \inf_{\mu \geq 0} \{(Ax - b)'\mu - c'x\}.$$

If $a_i'x - b_i < 0$ for any i, then $p(x) = -\infty$. Thus the dual of (D) is

$$\max -c'x \quad \text{or} \quad \min c'x$$

$$\text{subject to } A'x \geq b, x \geq 0$$

The Lagrangian optimality condition for (P) is

$$x^* = \arg\min_{x \geq 0} \left\{ \left(c - \sum_{i=1}^{m} \mu_i^* a_i \right)' x + \sum_{i=1}^{m} \mu_i^* b_i \right\},$$

from which we determine the complementary slackness conditions for (P):

$$\left(c_j - \sum_{i=1}^{m} \mu_i^* a_{ij} \right) x_j^* = 0, \quad x_j^* \geq 0, \quad \forall\ j = 1, \ldots, n,$$

$$c - \sum_{i=1}^{m} \mu_i^* a_i \geq 0, \quad \forall\ i.$$

The Lagrangian optimality condition for (D) is

$$\mu^* = \arg\min_{\mu \geq 0} \{(Ax^* - b)'\mu - c'x^*\},$$

from which we determine the complementary slackness conditions for (D):

$$Ax^* - b \geq 0,$$

$$(Ax^* - b)_i \mu_i^* = 0, \quad \forall\ i.$$

3.4.4 (www)

(a) Let λ_j be a Lagrange multiplier associated with the constraint $\sum_{i=1}^m x_{ij} = \beta_j$, and let ν_i be a Lagrange multiplier associated with the constraint $\sum_{j=1}^n x_{ij} = \alpha_i$. Define

$$X = \{x \mid x_{ij} \geq 0, \quad \forall\, i, j\}.$$

The Lagrangian function is

$$\begin{aligned} L(x,\nu,\lambda) &= \sum_{i,j} a_{ij}x_{ij} + \sum_{i=1}^m \nu_i \left(\alpha_i - \sum_{j=1}^n x_{ij}\right) + \sum_{j=1}^n \lambda_j \left(\beta_j - \sum_{i=1}^m x_{ij}\right) \\ &= \sum_{i,j}(a_{ij} - \nu_i - \lambda_j)x_{ij} + \sum_{i=1}^m \nu_i\alpha_i + \sum_{j=1}^n \lambda_j\beta_j. \end{aligned}$$

The dual function is

$$q(\nu,\lambda) = \inf_{x\in X} L(x,\nu,\lambda) = \begin{cases} \sum_{i=1}^m \nu_i\alpha_i + \sum_{j=1}^n \lambda_j\beta_j & \text{if } a_{ij} - \nu_i - \lambda_j \geq 0 \text{ for all } i, j, \\ -\infty & \text{otherwise.} \end{cases}$$

An alternative dual function is obtained by assigning a Lagrange multiplier λ_j to each constraint $\sum_{i=1}^m x_{ij} = \beta_j$, and lumping the remaining inequality constraints within the abstract set constraint. Thus,

$$X = \{x \mid \sum_{j=1}^n x_{ij} = \alpha_i, \quad x_{ij} \geq 0, \quad \forall\, i, j\}.$$

The Lagrangian function is

$$\begin{aligned} L(x,\lambda) &= \sum_{i,j} a_{ij}x_{ij} + \sum_{j=1}^n \lambda_j \left(\beta_j - \sum_{i=1}^m x_{ij}\right) \\ &= \sum_{i=1}^m \left(\sum_{j=1}^n (a_{ij} - \lambda_j)x_{ij}\right) + \sum_{j=1}^n \lambda_j\beta_j. \end{aligned}$$

Then the dual function is

$$\begin{aligned} q(\lambda) &= \inf_{x\in X} L(x,\lambda) \\ &= \sum_{j=1}^n \lambda_j\beta_j + \inf_{x\in X} \sum_{i=1}^m \left(\sum_{j=1}^n (a_{ij} - \lambda_j)x_{ij}\right) \\ &= \sum_{j=1}^n \lambda_j\beta_j + \sum_{i=1}^m \inf_{1\leq j\leq n} (a_{ij} - \lambda_j)\alpha_i, \end{aligned}$$

and the dual problem is

$$\begin{aligned} &\text{maximize} \;\; q(\lambda) \\ &\text{subject to} \;\; \lambda \in \Re^n. \end{aligned}$$

(b) & (c) The Lagrange multiplier λ_j can be interpreted as the price p_j. So if the transportation problem has an optimal solution x^*, then its dual also has an optimal solution, say p^*, and

$$q(p^*) = \sum_{i,j} a_{ij} x^*_{ij},$$

i.e.,

$$\sum_{j=1}^{n} p^*_j \beta_j + \sum_{i=1}^{m} \min_{1\le j\le n} (a_{ij} - p^*_j)\alpha_i = \sum_{i,j} a_{ij} x^*_{ij}. \tag{1}$$

Since x^* is primal feasible, we have

$$\sum_{j=1}^{n} p^*_j \beta_j = \sum_{j=1}^{n} p^*_j \sum_{i=1}^{m} x^*_{ij},$$

and by combining this with Eq. (1), we obtain

$$\sum_{i=1}^{m} \min_{1\le j\le n} \{a_{ij} - p^*_j\}\alpha_i = \sum_{i,j} (a_{ij} - p^*_j) x^*_{ij}. \tag{2}$$

By the feasibility of x^*, we have $\sum_{j=1}^{n} x^*_{ij} = \alpha_i$ for all i, and from Eq. (2) it follows that

$$\sum_{i,j} \big(a_{ij} - p^*_j - \min_{1\le j\le n} \{a_{ij} - p^*_j\}\big) x^*_{ij} = 0.$$

Since all the terms in the summation above are nonnegative, we must have

$$\left(a_{ij} - p^*_j - \min_{1\le j\le n} \{a_{ij} - p^*_j\}\right) x^*_{ij} = 0, \qquad \forall \ \ i, j.$$

Therefore if $x^*_{ij} > 0$, then

$$a_{ij} - p^*_j = \min_{1\le k\le n} \{a_{ik} - p^*_k\},$$

which can be equivalently expressed as

$$p^*_j - a_{ij} = \max_{1\le k\le n} \{p^*_k - a_{ik}\}.$$

Since p^* is arbitrary, this property holds for every dual optimal solution p^*.

3.4.5 (Duality and Zero Sum Games) www

Consider the linear program

$$\min_{\substack{\zeta e \ge A'x \\ \sum_{i=1}^{n} x_i = 1,\, x_i \ge 0}} \zeta,$$

whose optimal value is equal to $\min_{x\in X}\max_{z\in Z} x'Az$. Introduce dual variables $z \in \Re^m$ and $\xi \in \Re$, corresponding to the constraints $A'x - \zeta e \le 0$ and $\sum_{i=1}^n x_i = 1$, respectively. The dual function is

$$\begin{aligned} q(z,\xi) &= \inf_{x_i\ge 0,\, i=1,\ldots,n} \left\{ \zeta + z'(A'x - \zeta e) + \xi\left(1 - \sum_{i=1}^n x_i\right) \right\} \\ &= \inf_{x_i\ge 0,\, i=1,\ldots,n} \left\{ \zeta\left(1 - \sum_{j=1}^m z_j\right) + x'(Az - \xi e) + \xi \right\} \\ &= \begin{cases} \xi & \text{if } \sum_{j=1}^m z_j = 1,\ \xi e - Az \le 0, \\ -\infty & \text{otherwise.} \end{cases} \end{aligned}$$

Thus the dual problem, which is to maximize $q(z,\xi)$ subject to $z \ge 0$ and $\xi \in \Re$, is equivalent to the linear program

$$\max_{\xi e \le Az,\, z\in Z} \xi,$$

whose optimal value is equal to $\max_{z\in Z}\min_{x\in X} x'Az$.