A Simplifying Parametrisation of Maximal Cyclic Monotone Relations

M. Knott and C. S. Smith
London School of Economics and Political Science

Maximal Cyclic Monotone Relations

Dr M. Knott
Statistics Department
London School of Economics and Political Science
Houghton Street
London WC2A 2AE

Abstract

We investigate by new direct methods the structure of maximal monotone and maximal cyclic monotone relations. We obtain some new results for the case when cycle properties are limited to cycles of length no greater than m. We show that our approach gives easily several well known but deep results for maximal cyclic monotone relations. Our proofs are based on a geometric lemma of independent interest.
Keywords: Intersecting Spheres, Radial, Monge-Kantorovich

1 Introduction

An early paper by G. J. Minty [3] states his belief that 'some problems of mathematical physics that involve monotone operators could be better reformulated in terms of the graphs of the operators'. We have used this idea for the more restrictive cyclic monotone operators in a series of papers [8], [9], [2]. The importance of maximal cyclic monotone relations lies in their application to the Monge-Kantorovich problem. This seeks a distribution for $(X,Y) \in R^k \times R^k$ with given marginal distributions on R^k which minimises $\mathrm{E}\mid X - Y \mid^2$. A joint distribution concentrated on a maximal cyclic monotone relation will provide a solution to this problem.

In this paper we show how to use some of the ideas of Minty in [3] (also in [4], [5]) to obtain a useful simplifying property of maximal cyclic monotone relations (MCMR). The very simple underlying idea is that a MCMR Γ on $R^k \times R^k$ produces a one-to-one mapping if one rotates the corresponding pairs of axes from the two R^k through 45 degrees. This device was used by Minty to study maximal monotone relations, but surprisingly, when Rockafellar introduced the notion of cyclic monotonicity, he did not continue with it.

Section 2 gives the definition of MCMR and some background information. In Section 3 we give an intuitive result on the intersection of spheres in R^k. Section 3 contains the geometric result at the basis of the later work; in Section 4 writes the cycle property of a cyclic monotone relation in a form suitable application of the geometry. In Section 5 the geometry is applied to obtain some new results on cycle properties. The derivation of Rockafellar's representation of MCMR from ours is in Section 6. The last Section 7 quickly shows how to use our parametrisation to produce known results for radial transformations.

2 Cyclic Monotonicity

Throughout this paper, all the relations we consider are defined on $R^k \times R^k$, but we shall not in every case make this explicit.

A relation Γ on $R^k \times R^k$, ie a set of ordered pairs $\{(x,y),\, x \in R^k,\, y \in R^k\}$ is a *cyclic monotone relation of order* n (CMR[n]) if and only if for every sequence of pairs $(x_i, y_i) \in \Gamma$, $i = 1, \ldots, n$,

$$\sum_{i=1}^{n} x_i'(y_i - y_{i+1}) \geq 0, \tag{1}$$

where $y_{n+1} \equiv y_1$.

Notice that by choosing some of the pairs (x_i, y_i) the same it is easy to see that if Γ is a CMR[n] it is also a CMR[m] for positive integers $m < n$. A CMR[n] is a *maximal monotone cyclic relation of order* n (MCMR[n]) if no CMR[n] strictly includes it.

As a special case, a CMR[2] is known as a Monotone Relation (MR). A MCMR[2] is called a *Maximal Monotone Relation* (MMR).

Following Rockafellar [7], we define a *cyclic monotone relation* (CMR) as follows:

A relation Γ is a *cyclic monotone relation* (CMR) if and only if for every integer $n \geq 1$ Γ is a CMR[n].

A CMR is a *maximal cyclic monotone relation* (MCMR) if there is no CMR that strictly includes it.

Rockafellar characterises MCMR as the graphs of gradient functions of closed convex functions on R^k. His approach draws heavily on convex function theory. We prefer to investigate the MCMR more directly, and using geometrically intuitive methods to arrive at some of the deepest consequences of Rockafellar's characterisation, and at that characterisation itself. We have already given [2] an example of new results following from this approach in a slightly different area of application.

The next section gives an intuitive result which is the base for our later proofs and is of independent interest.

3 A geometric theorem

This result is a generalisation of a result first given by Kirzbraun referred to in [3].

Theorem 1 *Suppose that the set of closed spheres $\{T_i\}$ in R^k with square radii $\{r_i^2\}$ and centres $\{t_i\}$ have a non-empty intersection. A sufficient condition that a second set of closed spheres $\{U_i\}$ in R^k with square radii $\{r_i^2 + a_i^2\}$ and centres $\{u_i\}$ have also a non-empty intersection is that for all i, j*

$$| u_i - u_j |^2 \leq | t_i - t_j |^2 + a_i^2 + a_j^2. \tag{2}$$

The proof follows the approach of Valentine [10], who proved Kirzbraun's original result which took $a_i = 0$ for all i. The result is plausible because all pairs of spheres in the set $\{U_i\}$ have an intersection which is larger than that of the corresponding pair from the set $\{T_i\}$.

Following Valentine, it is sufficient to prove that every $(k + 1)$ subset of the spheres $\{U_i\}$, say $U_1, \ldots, U_{k+1}$ covers the simplex $\triangle$ generated by $u_1, \ldots, u_{k+1}$. The proof is by contradiction.

If $\cup_{i=1}^{k+1} U_i$ does not cover $\triangle$, then choose $t \in \cap_{i=1}^{k+1} T_i$ and $u \in \triangle - \cup_{i=1}^{k+1} U_i$. Then for all i

$$| u - u_i |^2 > | t - t_i |^2 + a_i^2. \tag{3}$$

From equation (2), for all i, j

$$| (u - u_i) - (u - u_j) |^2 \leq | (t - t_i) - (t - t_j) |^2 + a_i^2 + a_j^2$$

which implies using equation (3) that

$$(u - u_i)'(u - u_j) > (t - t_i)'(t - t_j).$$

Now we can follow Valentine's proof, because this is his equation (8). The contradiction follows from the assumption that $u \in \triangle$.

Corollary 2 *Given $\{u_i, t_i\} \in R^k \times R^k$ such that for all i, j equation (2) holds, it is possible for each given $t \in R^k$ to find $u \in R^k$ satisfying*

$$| u - u_i |^2 \leq | t - t_i |^2 + a_i^2. \tag{4}$$

This follows from Theorem 1. Construct spheres $\{T_i\}$ with centres t_i and radii $| t - t_i |$. These spheres clearly have point t in common. Construct spheres $\{U_i\}$ with centres u_i and square radii $| t - t_i |^2 + a_i^2$. It follows from Theorem 1 that the spheres $\{U_i\}$ have a point in common which can be taken as u.

4 Merging Two Cycles

Let us first write (1) for a sequence of $(m + n)$ pairs $(\tilde{x}_1, \tilde{y}_1), \ldots, (\tilde{x}_m, \tilde{y}_m)$, $(x_1, y_1), \ldots, (x_n, y_n)$. Simple algebra will show that for these pairs (1) may be written:

$$\begin{aligned} | (\tilde{x}_m - \tilde{y}_1) - (x_n - y_1) |^2 \quad \leq \quad & | (\tilde{x}_m + \tilde{y}_1) - (x_n + y_1) |^2 \\ & + 4 \sum_{i=1}^{m} \tilde{x}_i' (\tilde{y}_i - \tilde{y}_{i+1}) + 4 \sum_{i=1}^{n} x_i' (y_i - y_{i+1}), \end{aligned} \tag{5}$$

where $\tilde{y}_{m+1} \equiv \tilde{y}_1$ and $y_{n+1} \equiv y_1$.

In the special case $m = 1$, (5) simplifies to

$$| (\tilde{x}_1 - \tilde{y}_1) - (x_n - y_1) |^2 \leq | (\tilde{x}_1 + \tilde{y}_1) - (x_n + y_1) |^2 + 4 \sum_{i=1}^{n} x_i' (y_i - y_{i+1}). \tag{6}$$

There is a marked similarity between (5) and (2), and between (6) and (4). We use this in the next section to prove some fundamental properties of CMR$[n]$ and CMR.

5 Parametrisation of CMR[m]

In this section we show that maximality of a relation Γ is, in some cases, closely linked to the existence of $(x, y) \in \Gamma$ giving every possible value for $x + y$. The relation is more like a function if pairs of corresponding axes for x and y are rotated through 45 degrees.

Remark 3 *If Γ is a CMR[m] for a given $m \geq 2$, then for a given $t \in R^k$ there is at most one $(x, y) \in \Gamma$ with $x + y = t$.*

This is because if $t = x + y = \tilde{x} + \tilde{y}$ for $(x, y), (\tilde{x}, \tilde{y}) \in \Gamma$, (6) applied to the sequence of pairs $(\tilde{x}, \tilde{y}), (x, y)$ gives $\mid (\tilde{x} - \tilde{y}) - (x - y) \mid \leq 0$.

Theorem 4 *If for a given integer $m \geq 2$ Γ is a MCMR[m] and also a CMR[$2(m-1)$], then for each $t \in R^k$ one can find a unique $(x, y) \in \Gamma$ such that $t = x + y$.*

The uniqueness follows from remark 3 above.

The proof, by contradiction, that each t is represented in Γ uses Corollary 2. Suppose that for $\tilde{t} \in R^k$ there is no $(x, y) \in \Gamma$ such that $\tilde{t} = x + y$. We may consider whether it is possible to augment Γ with an additional ordered pair $(\tilde{x}, \tilde{y})$ with $\tilde{t} = \tilde{x} + \tilde{y}$ in such a way that the CMR[m] property is preserved. To preserve this property, for every collection $(x_1, y_1), \ldots, (x_m, y_m)$ of points in the augmented Γ, it is sufficient that (1) must be satisfied with $n = m$. If $(\tilde{x}, \tilde{y})$ appears in the collection r times, the sufficient condition is satisfied if $r = 0$. If $r \geq 1$, after simplifying to retain just one $(\tilde{x}, \tilde{y})$ and relabelling, the sufficient condition (1) becomes like (6) with $n = m - r$, the sum being interpreted as empty if $m = r$:

$$\mid (\tilde{x} - \tilde{y}) - (x_n - y_1) \mid^2 \leq \mid (\tilde{x} + \tilde{y}) - (x_n + y_1) \mid^2 + 4 \sum_{i=1}^{m-r} x_i'(y_i - y_{i+1}). \tag{7}$$

which can be written

$$\mid \tilde{u} - u_i \mid^2 \leq \mid \tilde{t} - t_i \mid^2 + a_i^2, \tag{8}$$

where $\tilde{u} = \tilde{x} - \tilde{y})$, $u_i = x_n - y_1$, $t_i = (x_n + y_1)$ and $a_i^2 = 4 \sum_{i=1}^{m-r} x_i'(y_i - y_{i+1})$. The last quantity is non-negative because Γ is a CMR[$m - r$]. It can be seen that (8) is of the same form as (4). The CMR[m] property will be preserved when Γ is augmented by $(\tilde{x}, \tilde{y})$ if (8) holds for all choices of $(x_1, y_1), \ldots, (x_m, y_m)$. This will be true if we can find $\tilde{u}$ satisfying all the restrictions (8). That this is so follows from Corollary 2, because since Γ is a CMR[$2n$], (7) guarantees that condition (2) holds.

Since Γ may be augmented by $(\tilde{x}, \tilde{y})$ with $\tilde{t} = \tilde{x} + \tilde{y}$ while remaining a CMR[m] it is not a MCMR[m], which contradiction completes the proof.

Immediate conclusions from Theorem 4 are:

Corollary 5 *If for a given integer $m \geq 2$ Γ is a MCMR[m] and also a CMR[$2(m-1)$], then Γ is MCMR[n] for $n = 2, \ldots, m$.*

Suppose that Γ can be augmented by $(x, y) \notin \Gamma$ while preserving it as a CMR[n], and that $t = x + y$. Then by Theorem 4 there exists $(\tilde{x}, \tilde{y}) \in \Gamma$ such that $t = \tilde{x} + \tilde{y}$. But then because the augmented Γ is a CMR[n], by remark 3, $(x, y) = (\tilde{x}, \tilde{y}) \in \Gamma$, so no augmentation is possible.

Corollary 6 *If Γ is a MMR, then for each $t \in R^k$ one can find a unique $(x, y) \in \Gamma$ such that $t = x + y$.*

This is an application of Theorem 4 with $m = 2$.

Corollary 7 *If Γ is a MCMR, then for each $t \in R^k$ one can find a unique $(x, y) \in \Gamma$ such that $t = x + y$.*

This follows from Theorem 4 because if Γ is a MCMR, then for some integer $m \geq 2$ it must be a MCMR[m], and it is also a CMR[$2(m-1)$].

Corollary 8 *If Γ is a MCMR, then it is MCMR[m] for all $m \geq 2$.*

If Γ is a MCMR, then for some integer $m \geq 2$ it must be a MCMR[m], and it is also a CMR[$2(m-1)$], so by Corollary 5 it is a MCMR[p] for $2 \leq p \leq m$. It is obviously a MCMR[p] for $p > m$.

It is easy to construct examples of MCMR[m]. For instance, the relations $\{(x, Ax), x \in R^2\}$ given by $A = \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}$ give a MMR which is not MCMR[3] for $a = 2$, and a MCMR[3] which is not MCMR[4] for $a = 1.5$, and a MCMR[4] which is not a MCMR[5] for $a = 1.25$. As a gets closer to 0, higher order cycle properties hold.

It is not clear whether Theorem 4 is the best possible result. One would wish to prove that if $m \geq 2$ there was a unique $(x, y) \in \Gamma$ with $x + y = t$ for Γ a MCMR[m]. This is stronger than Theorem 4,and implies that a MCMR[m] is also a MMR, but we do not know if it is correct.

6 Gradient Representations

In this section we show how to arrive at the known result that a MCMR is the (one to many) gradient mapping of a closed convex function from R^k to

R. We first show that the cycle properties (1) of a MCMR imply that a line integral (9) has a path independence property.

Suppose that Γ is a MCMR which by Theorem 4 may be written $\{(x(t), y(t)), t \in R^k\}$, where $(x(t), y(t))$ is the unique element of Γ with $x(t) + y(t) = t$. By Remark 3 $x(t)$ is continuous in t. For $t \in R^k$ let us define real valued functions $\alpha(t)$ and $\beta(t)$ by the line integrals

$$\begin{aligned} \alpha(t) &= \int_c x(u)'du \\ \beta(t) &= \int_c y(u)'du, \end{aligned} \tag{9}$$

where c is a smooth path from 0 to t.

The function $\alpha(t)$ of (9) does not depend on the particular path c between 0 and t. One can see that a line integral of the same differential form over a closed smooth path of finite length takes the value 0, because one can consider it as approximated by

$$\sum_{i=1}^{n} x(u_i)'(u_i - u_{i+1}), \tag{10}$$

where the u_i are points on the path and $u_1 = u_{n+1}$. Since $u_i = x(u_i) + y(u_i)$ the cycle property (1) implies that

$$\sum_{i=1}^{n} x(u_i)'(u_i - u_{i+1}) \geq 0 \geq \sum_{i=1}^{n} x(u_{i-1})'(u_i - u_{i+1}). \tag{11}$$

It follows from (11) that

$$\begin{aligned} \mid \sum_{i=1}^{n} x(u_i)'(u_i - u_{i+1}) \mid &\leq \mid \sum_{i=1}^{n} (x(u_i) - x(u_{i-1}))'(u_i - u_{i+1}) \mid \\ &\leq \sum_{i=1}^{n} \mid x(u_i) - x(u_{i-1}) \mid\mid u_i - u_{i+1} \mid . \end{aligned}$$

The last expression may be made as close to zero as one wants by choosing the u_i sufficiently close together to make all the $\mid x(u_i) - x(u_{i-1}) \mid$ small, remembering that $\sum_{i=1}^{n} \mid u_i - u_{i+1} \mid$ is bounded above by the length of the path. It follows that (10) must tend to zero as it approximates the integral over the closed path more closely.

In the same way, the function $\beta(t)$ of (9) does not depend on the particular path c between 0 and t. The functions $\alpha(t)$ and $\beta(t)$ are therefore differentiable, with $d\alpha(t)/dt = x(t)$ and $d\beta(t)/dt = y(t)$. It is clear also that they are they are convex, because an approximation for $\alpha(t)$ can be written

$$\sum_{i=2}^{n} x(u_i)'(u_i - u_{i-1}), \tag{12}$$

where the u_i are points on the path from $u_1 = 0$ to $u_n = t$. Any such approximation is $\geq x(u_1)'(u_n - u_1)$ by (1), which shows convexity of $\alpha(t)$. Similarly for $\beta(t)$.

It follows from (9) that

$$\begin{aligned} \alpha(t) + \beta(t) &= \int_c (x(u) + y(u))' du \\ &= \int_c u' du \\ &= \mid t \mid^2 /2. \end{aligned} \tag{13}$$

The representation of a MCMR Γ in the form $\left\{ \frac{d\alpha(t)}{dt}, \frac{d\beta(t)}{dt} \mid t \in R \right\}$ is, we believe, superior to the Rockafellar representation of Γ as a one-many function based on the subgradients of a convex function, say λ. We can get Rockafellar's convex function λ by using conjugate functions

$$\begin{aligned} \alpha^*(x) &= \sup_t (x't - \alpha(t)) \\ \beta^*(y) &= \sup_t (y't - \beta(t)). \end{aligned} \tag{14}$$

Note that if $(\tilde{x}, \tilde{y}) \in \Gamma$ and $\tilde{x} + \tilde{y} = \tilde{t}$, then because $x't - \alpha(t)$ and $y't - \beta(t)$ are concave differentiable functions of t, from (14)

$$\begin{aligned} \alpha^*(\tilde{x}) &= \tilde{x}'\tilde{t} - \alpha(\tilde{t}) \\ \beta^*(\tilde{y}) &= \tilde{y}'\tilde{t} - \beta(\tilde{t}). \end{aligned}$$

and so

$$\alpha^*(\tilde{x}) + \beta^*(\tilde{y}) = \mid \tilde{x} + \tilde{y} \mid^2 . \tag{15}$$

If x does not appear in any $(x, y) \in \Gamma$, then $\alpha^*(x) = \infty$, and similarly $\beta^*(y) = \infty$ if y does not appear in any $(x, y) \in \Gamma$. Directly from the definitions (14), for all t,

$$\alpha^*(x) + \beta^*(y) \geq -\frac{1}{2} \mid t - (x + y) \mid^2 + \frac{1}{2} \mid x + y \mid^2, \tag{16}$$

and taking $t = x + y$,

$$\alpha^*(x) + \beta^*(y) \geq \frac{1}{2} \mid x + y \mid^2 . \tag{17}$$

From (15), (17) and the remark just before (16) it follows that for all x and for $(\tilde{x}, \tilde{y}) \in \Gamma$

$$\alpha^*(x) - \frac{1}{2}x^2 - (\alpha^*(\tilde{x}) - \frac{1}{2}\tilde{x}^2) \geq (x - \tilde{x})'\tilde{y}. \tag{18}$$

It follows immediately from (18) that $\lambda(x) = \alpha^*(x) - \frac{1}{2}x^2$ is a closed convex function with subgradient $\tilde{y}$ at $\tilde{x}$, and so gives the Rockafellar representation of Γ.

7 Radial Relations

It is easy to apply the methods of this paper to obtain the results on cyclic monotonicity of radial transformations due to Cuesta, Rüschendorf and Tuero-Diaz [1]. Suppose that $a(s)$ and $b(s)$ are two differentiable non-decreasing convex functions on R^+ (with derivatives $a_1(s)$, and $b_1(s)$) such that

$$a(s) + b(s) = \frac{1}{2} \mid s \mid^2 . \tag{19}$$

$a_1(0) = b_1(0) = 0$. Then if $t \in R^k$, we have differentiable convex functions $\alpha(t) = a(\mid t \mid)$, and $\beta(t) = b(\mid t \mid)$ that give a MCMR

$$\Gamma = \left\{ \frac{d\alpha(t)}{dt}, \frac{d\beta(t)}{dt} \mid t \in R^k \right\}.$$

It follows immediately that

$$\Gamma = \left\{ (a_1(\mid t \mid)t/ \mid t \mid, b_1(\mid t \mid)t/ \mid t \mid) : t \in R^k \right\}, \tag{20}$$

where $t = 0$ is taken as giving the pair $(0, 0)$.

The form (20) can be written

$$\begin{aligned} \Gamma &= \left\{ (x(t), x(t) b_1(\mid t \mid)/a_1(\mid t \mid)) : t \in R^k \right\} \qquad (21) \\ &= \left\{ (x(t), x(t)(\mid t \mid - a_1(\mid t \mid)/a_1(\mid t \mid)) : t \in R^k \right\}. \end{aligned}$$

Defining an inverse function a_1^{-1} on R^+ by $a_1^{-1}(s) = \inf_w \{w \mid a_1(w) \geq s\}$, we can construct a function $\theta : R \to R$ by

$$\theta(s) = a_1^{-1}(s) - s,$$

with $\theta(0)$ taken as 0.

Then

$$\Gamma = \left\{ (x(t), \theta(\mid x(t) \mid) x(t)/ \mid x(t) \mid) : t \in R^k \right\}, \tag{22}$$

This gives Proposition 3.1 of [1], because one can obtain any monotonically non-decreasing function θ by a suitable choice of the functions a and b. In fact, if θ is non-decreasing on R^+ with $\theta(0) = 0$, then one can define

$$\phi(t) = inf_s \{s \mid s + \theta(s) \geq t\}. \tag{23}$$

Then $\phi(t)$ and $t - \phi(t)$ are both non-decreasing functions on $t \in R^+$, and one may take

$$\begin{aligned} a(t) &= \int_0^t \phi(u) du \qquad (24) \\ b(t) &= \int_0^t (u - \phi(u)) du. \end{aligned}$$

8 Conclusions

In [6] the author discusses the theorem about the maximal monotonicity of subdifferentiable mappings, and remarks:

> Most books which use this theorem avoid giving a proof (referring the reader to the second of Rockafellar's three proofs) ...; at least one book reproduces Rockafellar's flawed second proof. (This fundamental theorem still awaits a really simple proof.)

Perhaps our approach provides such a simple proof.

The extension of our results to Hilbert Space is straightforward, though it would need care to set them in a Banach space.

Two lines of further work are now under way. We are looking more carefully at the cycle properties of linear operators, and we are trying to extend the ideas in this paper to relations between more than two sets of vectors as in [2].

References

[1] J. A. Cuesta, L. Rüschendorf, and A. Tuero-Diaz. Optimal coupling of multivariate distributions and stochastic processes. *Journal of Multivariate Analysis*, 46:355–361, 1993.

[2] M. Knott and C. S. Smith. On a generalisation of cyclic monotonicity and distances among random vectors. *Journal of Linear Algebra and its Applications*, 199:363–371, 1994.

[3] G. J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29:342–346, 1962.

[4] J.-J. Moreau. Fonctions convexes duales et points proximaux dans une espace hilbertien. *Comptes Rendus de l'Académie des Sciences de Paris*, 255:2897–2899, 1962.

[5] J.-J. Moreau. Propriétés des applications ≪ prox ≫. *Comptes Rendus de l'Académie des Sciences de Paris*, 256:1069–1071, 1963.

[6] R. R. Phelps. *Convex Functions, Monotone Operators and Differentiability*. Springer, Berlin, 1989.

[7] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[8] C. S. Smith and M. Knott. A note on the optimal transportation of distributions. *Journal of Optimization Theory and its Applications*, 52:323–329, 1987.

[9] C. S. Smith and M. Knott. On Hoeffding-Fréchet bounds and cyclic monotone relations. *Journal of Multivariate Analysis*, 40:328–334, 1992.

[10] F. A. Valentine. A Lipschitz condition preserving extension for a vector function. *American Journal of Mathematics*, 67:83–93, 1945.